AD-A280 182

Scientific Research Associates, inc.

50 Nye Road, P.O. Box 1058
Tel: (203) 659-0333

Glastonbury, Connecticut 06033
Fax: (203) 633-0676

REPRINTS

THE PHYSICS AND OPERATION OF ULTRA-SUBMICRON LENGTH SEMICONDUCTOR DEVICES

DTIC
ELECTE
JUN 09 1994
S G D

Submitted to
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

DTIC QUALITY INSPECTED 2

May 1994

Approved for Public Release;
Distribution Unlimited

94-17516

94 6 8 122

January 23, 1994

Modeling of Quantum Transport in Semiconductor Devices

David K. Ferry
Center for Solid State Electronics
Arizona State University
Tempe, AZ 85287-6206

Harold L. Grubin
Scientific Research Associates
P. O. Box 1058
Glastonbury, CT 06033

Accesion For
NTIS CRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution /
Availability Codes
Dist A-1
Avail and / or Special

Abstract

The evolutionary decrease in the size of an individual semiconducto device continues with no apparent end of the process in sight. As a consequence, it is quite likely that critical dimensions will soon be comparable to quantum coherence lengths for the particles involved in the transport within the device. Generally, quantum transport differs from semi-classical transport in the utilization of a quantum kinetic equation (as opposed to the Boltzmann transport equation). These quantum kinetic equations can be developed for the density matrix, the Wigner distribution function, and real-time Green's functions, as well as for many reduced approximations to these quantities. In this review, we study how these various approaches are connected as well as how they offer different views into the quantum behavior within devices. Considerable attention is given to tunneling heterostructures

and the resonant-tunneling diode, as well as to the quantum dot structure, which is the single-electron limit of latter device. An attempt is made to also identify those areas which warrant further investigation as well as to review what has been accomplished in the field.

Contents

I Introduction 4
1 Quantization in Devices 6
2 The Differences from Boltzmann Transport 8
A Statistical Thermodynamics and Quantum Potentials 10
B Phase Interference 13
3 Open Systems and Contacts 16
A Ballistic Transport 17
B Role of the Boundaries and Contacts 19
4 Some Potentially Important Quantum Devices 21
A The Resonant-Tunneling Diode 22
B Quantum Dots 23

II The Quantum Equations 24
5 The Density Matrix and Its Brethren 28
A The Liouville and Bloch Equations 30
B Wigner Functions and Green's Functions 31
C Reduced Density Matrices and Projection Operators . 36
6 The Kubo Formula and Langevin Equations 39
A The Kubo Formula and Correlation Functions 39
B Retarded Langevin Equations 42
7 Boltzmann-Like Approaches 45
8 Moment Equations—The Transition from Classical to Quantum 46
A The Hydrodynamic Moment Equations 49
B Applications in Modeling Devices 52

III Modeling with the Density Matrix 54
9 Some Considerations on the Density Matrix 55
A Statistics of a Single Barrier 58
B Multiple-Barrier Structures 60
10 Dissipation and Current Flow 62
11 Further Considerations on the Density Matrix 68

A An Alternative Approach to the Density Matrix . . . 68
B Differential Capacitance 72

IV Modeling with the Wigner Distribution 74
12 Methods of Solving the Equations 76
A The Initial State 76
B Numerical Discretization and Solutions 79
C Boundary Conditions for the Simulation 82
13 The Double-Barrier Rescnant-Tunneling Diode 84
14 The Role of Dissipation 88
15 Other Devices 90

V Modeling with the Green's Functions 91
16 Homogeneous, Low-Field Systems 93
A The Retarded Function 95
B The "Less-Than" Function 99
17 Homogeneous, High-Field Systems 104
A The Airy Function Retarded Green's Function 106
B The Less-Than Function 110
18 Femtosecond Laser Excitation 113
19 The Green-Kubo Formula 118
20 The Resonant-Tunneling Diode 124

VI Acknowledgements 126

I. Introduction

Since the introduction of integrated circuits, the number of individual transistors on a single chip has doubled approximately every three years. Today, we are looking at multi-megabit dynamic random access memories (the 16 Mb is on the market, the 64 Mb is in preproduction and commercial sales are expected in 1995, and the 256 Mb has already appeared in research versions). Comparable densities of transistors, our prototypical semiconductor device, are achieved in dense signal-processing chips, and microprocessors are only slightly less dense. The annual progression of the increase of device density has followed a well-developed set of scaling laws for at least the last two decades,[1,2] and there is no indication of any deviation from this scaled progression for the next decade or so. At the rate of progress of dynamic memory, we can expect to reach chip densities of 10^9 devices by 2001. By the year 2020, we may well have memory chips with a density of 1 terabit, providing a number of interconnection and architectural problems can be overcome. Terabit memory chips imply a quite small transistor, and in fact, the scaling rules mentioned above imply a certain reduction of design rule (which is reflected in gate length and metal line width). The reduction commensurate with the present growth of integration density is approximately a factor of 1.4 in gate length for each new device generation (which produces only an increase of 2× in density, the remainder coming from circuit enhancements and larger chip size). This means that we will be using $0.1 - 0.15$ μm design rules for the 4 Gb chips around 2005. If we continue this extrapolation, current technology will dictate reduction to 30 nm design rules, and a cell size below 10^3 nm^2, for the 1 terabit memory.

Whether or not the above scaling rules continue to hold, it appears that we will eventually see devices with gate lengths of 50 nm and below as part of real integrated circuits. An electron traveling at the saturated velocity (of most semiconductors) will traverse this length in about 0.5 ps, or approximately the time duration of the transient response of an electron to an instanteously applied electric field of 50 kV/cm. Moreover, the inelastic mean free path (the distance over which the carriers travel between energy dissipating scattering processes, or over which they lose quantum mechanical phase information) is about 0.1 μm (slightly smaller in Si). This is greater than the gate length expected in these small devices. Thus, it is expected that quantum effects will become quite significant in the operation of such devices.

While very few laboratories have made research devices on a 0.1 μm scale, there is evidence from ultra-submicron devices that have been made that quantum effects will be important. Silicon MOSFETs (metal-oxide-semiconductor field-effect transistors) have been made with gate lengths as short as 60-70 nm.[3,4] GaAs Schottky-gate FETs with gates as short as 30 nm[5-7] and high-electron mobility transistors with gate lengths as short as 20 nm[8,9] have been made. In the shortest of these research devices, there is clear evidence that tunneling, a quantum mechanical effect, through the gate depletion barrier is the dominant contributor to the current control, which much reduces the gate control of the current.[10]

The transport of carriers in semiconductor devices has long been a subject of much interest, not only for material evaluation, but also in the realm of device modeling and, more importantly, as an illuminating tool for delving into the physics governing the interaction of electrons (and/or holes) with their environment.[11] Moreover, the careful modeling of transport and interactions in devices allows one to push the technology to ever smaller devices successfully, accounting for new effects arising from the smaller sizes.[4]

From the above discussion, it appears that more detailed modeling of quantum contributions needs to be included in device modeling for future ultra-small devices.[12] These quantum effects appear in many guises: a) modification of the statistical thermodynamics within the device (and in its connection to the external world), b) introduction of new length scales, c) ballistic transport and quantum interference, and d) new fluctuations affecting device performance. Many of these effects already have been studied, either in models of ultra-submicron devices or in macroscopic devices at low temperatures (which, more appropriately, may be referred to as *structures*, since they may well not be true devices in the normal sense). In this review, we first will try to emphasize the nature of these differences and some of the new effects and review what is known about them. Then, we will try to put the approaches to quantum transport for devices into context with each other. Finally, we will review the manner in which each approach has been used to model several prototypical quantum devices. We will not review the entire field of mesoscopic devices, which have been studied extensively at low temperatures, as these have been the subject of several excellent reviews in recent years.[13–15]

1. Quantization in Devices

Today, for the greater part of device and circuit design, relatively simple device and circuit models, equivalent circuits as it were, are used.[16] This type of approach has been integrated into VLSI design codes, and into microwave use for discrete systems as well. In both cases, the results have been quite good *for today's devices.* These equivalent circuit models are based largely upon quite simplified transport analysis for the carriers within the device. Nevertheless, many more quite sophisticated models of transport, universally based upon the Boltzmann equation,[11] are used to evaluate devices in the submicron and ultra-submicron regime.[17] The observations of velocity overshoot, in which the transient dynamic response of the carriers becomes important in the device performance, has been a primary driver for using the more complicated, and more physically correct, transport models.[10]

In detailed modeling of semiconductor devices, one normally couples a more-or-less detailed transport model with a solution of the Poisson's equation for the specific structure being modeled.[18] Quantization can appear in either of these two basic parts of the device model. It has been known for a great many years that carriers in the inversion layer of a Si MOSFET are confined by the barrier between the semiconductor-oxide interface on one side and the band bending of the conduction band on the other side. Since the average thickness of the inversion layer is comparable to the de Broglie wavelength of the electrons, this confinement is sufficient to produce quantization in the direction normal to the oxide-semiconductor interface.[19] By constraining the motion normal to this interface, the carrier motion is now allowed only in the two directions parallel to the interface, and a *quasi-two-dimensional electron (or hole) gas* is formed. In the case of Si, the six-fold degenerate valleys of the conduction band are split, with the two valleys having the heavy longitudinal mass normal to the interface lying lower in energy than the remaining four valleys having the light transverse mass normal to the interface. This quantization is important in determining the number of carriers in the inversion layer, and appears as an extra contribution to the gate capacitance—the peak of the wave function lies away from the interface, near the center of the quantum well formed by the barriers on either side, which is different from the classical case where the density peaks at the oxide-semiconductor interface. This quantization is also seen in the high-electron mobility transistor, or HEMT,[20–24] and is also important for transport in quantum wells,[25] and for the detailed screening of the carriers in these structures.[26] Consequently, quite complicated simulation codes have been developed to accurately determine the wave functions and charge density self-consistently in the quantized inversion layer.[27,28]

The transport of the carriers along the channel in the above-mentioned devices is still usually treated by semi-classical techniques—primarily through studies based upon the Boltzmann transport equation. In these approaches, classical transport physics is used with the scattering processes calculated from quantum mechanical approaches, usually no more complicated than the Fermi golden rule.[11] However, the onset of quantum mechanical problems in transport has been the subject of considerable discussion.[12] In general, the semi-classical approach assumes that the scattering processes are perturbations distinct from those of the driving fields, that the scattering occurs instanteously in both space and time, and that potential and density gradients are slow on the scale of the de Broglie wavelength of the carriers.[29] In future ultra-submicron semiconductor devices, all of these assumptions can be expected to be violated. Some work has already appeared concerning the interaction of the driving fields and the scattering processes, an effect known as the intra-collisional field effect (ICFE).[30–34] The problem of the rapid spatial variation of the potential is of course what leads to the quantization effects in the first place,[35] and the multiple interactions this causes is the major problem to be addressed in this review. The transition between semi-classical dynamics and quantum dynamics is one that remains in question in basic quantum theory,[36] but quantum transport has been discussed for some time. One aspect of this is that the basic equations are Markovian in nature, but under the conditions in which a one-electron distribution function is used, these can become non-Markovian in nature due to memory effects introduced by the scattering. Under strong fields and scattering, a new non-perturbative basis of electron states, rather than a simple perturbation of the Boltzmann equation, needs to be used.[37] To be sure, this problem—the steady-state of the far-from-equilibrium system under high fields—is not new, and appears equally as well in the semi-classical transport problem. The first to suggest this new dissipative steady-state was different was Landauer.[38] *It was pursued extensively by the Brussels group,[39] but the major point above is that the transition from semi-classical to quantum dynamics is also not a simple perturbative process. It is these major differences that create much of the problem in trying to develop quantum mechanical treatments of the transport for strongly non-equilibrium systems such as occur in semiconductor devices.*

2. The Differences from Boltzmann Transport

The basic transport equation for studying carrier behavior in semi-classical models of semiconductor devices has been the Boltzmann transport equation:

$$\frac{\partial f}{\partial t} + \mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{r}} + e\mathbf{F}\cdot\frac{\partial f}{\partial \mathbf{p}} = \sum_{\mathbf{p}'}[S(\mathbf{p},\mathbf{p}')f(\mathbf{p}') - S(\mathbf{p}',\mathbf{p})f(\mathbf{p})] \qquad (1)$$

where $\mathbf{p} = m\mathbf{v}$ is the momentum, and $\mathbf{r}$ is the position. Here, it is assumed that the momentum is related to the energy of the particles by a well-defined single-electron band structure; e.g., the *spectral density* is defined by $A(E,\mathbf{p}) = \delta(E - p^2/2m)$ (the spectral density is related to the dispersion relation between energy and momentum; integration over the vector momentum produces the density of states). Moreover, it is also assumed that the effect of the potential arises solely from the value of the first derivative, the field $\mathbf{F}$ in (1). Finally, it is assumed that the distribution function $f(\mathbf{p},t)$ varies slowly on the temporal scale of the relaxation processes, in that it is the local distribution at time t that appears in (1) and not some retarded value of the distribution (which one would assume would be the distribution at the time the appropriate free path began). There are then two approaches to solving this equation to obtain transport coefficients:

- It is assumed that the variation of the distribution from the equilibrium Maxwell-Boltzmann one (nondegenerate statistics are assumed) is small, and the value of f in the derivatives is replaced by the equilibrium value. This leads to what is usually referred to as the relaxation-time approximation.

- For complicated, anisotropic scattering processes, or for high-field transport, the above approximation fails, and one must actually solve for the distribution function. This, in fact, is the major problem in hot carrier transport.

In the case of quantum transport, each of the above assumptions fails. In particular, the spectral density is no longer a simple delta function, and one must find its form in the interacting system of many electrons with scattering by impurities, phonons, and other electrons. At low temperatures, and near equilibrium, the spectral function is usually found to be a Lorentzian, in which broadening exists around the value specified for the energy-momentum relation of the semi-classical model. In addition, the potential leads to non-local behavior, in which the last term on the left-hand side of (1) includes an entire heirachy of derivatives, such as originally introduced by Wigner.[40]

Finally, the collisions are no longer localized in space and time, so that the collision integral on the right-hand side of (1) becomes a non-Markovian retardation integral. This leads to a modified heirarchy of solutions for the quantum distribution function:

- The spectral function must first be determined in the interacting system.
- For near-equilibrium systems, or at low temperatures, it may be assumed that the quantum distribution is given by small deviations from the equilibrium Fermi-Dirac distribution.
- For complicated, anisotropic scattering processes, or for high-field and strongly non-equilibrium transport, the above approximation fails, and one must actually solve for the distribution function.

To be sure, in some approaches this sequence is finessed by using single-time functions, such as the density matrix and the Wigner distribution function, which essentially integrate out the spectral function, but retain the full spatially non-local nature of the potential interactions that lead to the heirarchy of derivatives appearing in the transport equation which replaces (1). We will illustrate this further below. Nevertheless, the transport problem, as in the semi-classical case, remains a balance between the driving forces, primarily the potential, and the relaxation forces represented in the collision integral.[41]

In the quantum mechanical case, there has been an argument for some time over whether or not the application of an electric field to a crystal would destroy the bulk band structure and create a Stark ladder of discrete states.[42] In fact, it is known that this does not occur in bulk crystals, where the use of the electric field creates a Franz-Keldysh shift of the bands, which is quite useful in modulated electroreflectance to study the band structure.[43] Some Stark ladder effects are seen in well correlated superlattice structures under optical illumination, but, in general, the effects are washed out in bulk materials by the scattering processes found there.[44]

A. Statistical Thermodynamics and Quantum Potentials

As we discussed above, the potential in quantum systems creates actions that are nonlocal to the actual potential, i.e., they can occur some distance from the potential. Let us consider how this nonlocality arises. Consider a simple potential energy barrier (Fig. 1) $V(x) = eV_o u(-x)$, where $u(x)$ is the Heavyside step function. We assume that a non-zero density exists in the region $x > 0$, and the question is how the density varies near the barrier, a quite typical problem in introductory quantum mechanics and in devices. Here, however, the problem refers in general to a statistical mixed state, rather than to a single quantum state. In classical mechanics (in the absence of any self-consistent Poisson equation solutions to find a new, self-consistent potential), the density varies as $\exp(-\beta V)$, where β is the inverse electron temperature ($= 1/k_B T$), and for this case is uniform and constant up to the barrier, dropping abruptly to zero in the half-space $x < 0$. In quantum mechanics, however, the wave function is continuous, and for any member of the statistical ensemble must be small at the interface (vanishingly small for the case $V_o \to \infty$). This then leads to a different behavior on the part of the density. In Fig. 1, we show the Wigner distribution function (which, for the moment, can be thought of as the quantum statistical mechanical analog of the classical phase-space distribution) for this situation. The parameters here are appropriate to bulk GaAs, with $n = 2 \times 10^{17}$ cm^{-3}.[45] We note that, far from the barrier, the distribution approaches the classical Maxwellian form, but near the barrier, the distribution differs greatly from the uniform classical case. The repulsion of density from the barrier is required by the vanishing of the wave function at the barrier, but the first peak in the wave function away from the barrier occurs closer to the barrier for higher momentum states. This leads to much of the complication evident in the figure, and to a momentum-dependent positional correction to the density away from the potential barrier. The density peak away from the barrier is governed by physics similar to the peak in density away from the semiconductor-oxide interface in a MOSFET, and assures that net charge neutrality is maintained (which means that Poisson's equation is included in the solution). The deviation in the density occurs over several *thermal* de broglie wavelengths (evaluated with the thermal momentum) $\lambda_D = \sqrt{\hbar^2/3mk_B T}$. This suggests that nonlocal deviations from classical results can be expected to occur in most semiconductor devices over a range of 20-40 nm even at room temperature!

It is clear that the density no longer varies simply as $\exp(-\beta V)$, and that modifications to the statistical mechanics need to be made. The development of quantum corrections to statistical thermodynamics, especially in equilibrium, has a rich and relatively old history. Unfortunately, there is no consensus as to the form of the correction to this simple exponential behavior. If we could find such a correction, it could be utilized in the semi-classical hydrodynamic equations developed from moments of more basic transport equations such as (1).

One of the original efforts to obtain quantum corrections to classical distribution functions was done by Wigner, in introducing the Wigner distribution function.[40] In this regard, it can be considered as an attempt to find an additional term that can be added to the classical potential to produce the desired results. The Wigner potential has been put in the form[46]

$$U_W = -\frac{\hbar^2}{8m}\frac{\partial^2(\ln n)}{\partial x^2}. \tag{2}$$

This represents a quantum correction to the mean kinetic energy of a distribution of particles. Bohm[47] also introduced an effective potential, in his discussions. For a distribution of particles, in the single-electron approximation, the Bohm potential represents a non-electrostatic force, acting upon a particle distribution whose value is determined by the *form* of the particle distribution. In a sense, this potenial is determined through an interaction of the particle with itself quantum mechanically. The Bohm potential is given by

$$U_B = -\frac{\hbar^2}{2m\sqrt{n}}\frac{\partial^2\sqrt{n}}{\partial x^2}. \tag{3}$$

These two differ numerically only in a minor way, even though their conceptual origins are quite different.

Feynman and Hibbs[48] suggested a variational approach by which the classical potential would be weighted by a Gaussian spreading function. Later work by Feynman and Kleinert[49] extended this to the development of a general variational form for the effective potential, in which a nonlocal smoothing function is applied to the actual potential, and new terms arise to represent quantum diffusion. A new version, based upon a Green's function solution of the effective Bloch equation for the density matrix in the nonlocal potential has been developed, but untried in actual device simulations.[50] This will be discussed further below.

It must be emphasized that the variation of the wave functions, or the consequent quantum distribution functions, away from confining barriers leads to quantization within a small system. This quantization is the over-riding property of small systems, such as quantum wires and quantum boxes.[51] In some cases, the narrow minibands that result from this quantization have been suggested as a method of cutting down on phonon scattering, by insuring that the width of a miniband is small compared to the optical phonon energy, while the spacing of the minibands is larger than this energy.[52] Most devices, however, are (erroneously) thought to be unconstrained in the direction to/from the contacts (or the reservoirs, as they will often be called below), so that these effects are not likely to be observed in most realistic devices. In fact, determining the contact effects in these "open" systems will be quite complicated, a point to which we return below.

B. Phase Interference

While current devices have gate lengths in the 0.25-0.7 μm range, at least for production devices (the shorter ones are GaAs microwave devices), future devices will reach far smaller sizes. It is conceivable that the gate length will then be comparable to those in which quantum effects are studied. The relevant quantity for discussion of quantum interference effects is the phase change of the carrier as it moves through the semiconductor device. Interference between differing electron waves, or differing electrons on their individual trajectories, can occur over distances on the order of the coherence length of the carrier wave, and this latter distance is generally taken to be the inelastic mean free path, or phase breaking length. Ballistic, and therefore coherent and unscattered, transport has been observed through the base region of a GaAs/AlGaAs hot electron transistor.[53] From this, it is estimated that the inelastic mean free path for electrons in GaAs may be as much as 0.12 μm *at room temperature.* The phase-breaking length, or inelastic mean free path, is of the order of (and usually equal to) the energy relaxation length $l_e = v\tau_e$, where τ_e is the energy relaxation time and v is a characteristic velocity. This tells us that even in Si the electron inelastic mean free path may be 50–100 nm. Thus, the inelastic mean free path can be quite long, and can be comparable to the gate length in these devices. Since the phase remains coherent over the range of the correlation function of the electrons (in space or time), there can be interference effects in the overall conductance of the device. The small device will then reflect the intimate details of the impurity distribution in the particular device, and macroscopic variations can then arise from one device to another, an effect well known in mesoscopic devices, where it leads to nonlinearities and fluctuations.[54] The basic concepts were expressed rather early by Landauer,[55] in which the conductance through a region with localized scatterers was expressed by very sample specific properties, known as the Landauer formula

$$G = \frac{2e^2}{h} T. \tag{4}$$

Here, the formula is expressed for one dimension, but it can be expanded to more dimensions by interpreting T as the total transmission of all modes (electrons) through the region of interest. In this formula, the potentials, used to calculate the conductance, are determined at the reservoirs (or contacts). While the original formula was obtained for noninteracting electrons, recent work has shown that a similar, but more complicated, form is obtained in the interacting electron case.[56]

The most usual study of the sample specific variations of the conductance with gate bias, applied bias, or magnetic field, all of which provide fluctuations in the local potential in the inhomogeneous sample (and all samples are inhomogeneous in this phase coherent regime), has dealt with universal conductance fluctuations.[15] However, it is also possible to have a net coherent backscattering from the impurities, without losing the phase coherence, and this leads to the concept of weak localization, a form of increased resistance due to the interactions.[15,57] One additional effect which has been suggested, but not studied well, is the fact that the random impurities cause a significant deviation in the current density from the uniform (average) value, especially where the cross-section of a single scatterer exceeds that of its equivalent volume of background semiconductor. Then, any single scatterer is likely to affect a greater current, due to the detour of current lines away from other scatterers. This can lead to a greater effect of each scatterer and, hence, a larger contribution to the resistance of the device.[58]

The above effect reaches its pinacle in the presence of a magnetic field, which can be coupled through the two phase coherent paths, and which leads to the Aharonov-Bohm effect.[59] The effect is most commonly studied in metal loops, coupled to a pair of reservoirs.[60–64] However, anytime two mutually uncorrelated quantum channels are connected at the reservoirs, one must expect that there will be flux sensitive fluctuations of the conductance.[65–70] One can ask the question as to whether the presence of inelastic scattering in one or both arms of the loop will cause these oscillations to be damped, and the answer is generally in the positive. The study of this damping effect led to the general development of the multi-channel version of (4).[71] Nevertheless, the presence of many channels of transport through the active gate region can be expected to lead to relatively large fluctuations in the overall conductance, if the conditions are properly attained.

Each of these effects is likely to begin to impact devices, as the size is reduced, even at room temperature. The most likely is universal conductance fluctuations, especially in the turn-on characteristics of the device, where the conductance is low and the impurities are being charged/discharged. Consider a small device, perhaps with a gate length and width of 0.05×0.1 μm, respectively. If the number of carriers in the inversion channel is 2×10^{12} cm^{-2}, there are only 100 electrons under the gate. If there is a fluctuation of a single impurity (between ionized and neutral), one might expect a change of conductance of order 1% in the thermally averaged classical regime. The change can be much larger if the carriers are phase coherent. The phase coherence[72] and the charging of such single impurities has been detected at low temperature in Si MOSFETs.[73] It is clearly established now that the effect can be much larger than one would expect, and this largeness is due to the quantum interference caused by the change in trajectories of individual electrons. The conductance change in the phase interference process can be of the order of (4), which is about 40 μS. If our device were to exhibit outstanding conductance of 1000 mS/mm (of gate width), the absolute conductance would only be 100 μS, so that the fluctuation could be of the order of 40% of the absolute conductance! This is a very significant fluctuation, and arises from the lack of ensemble averaging of these effects in the phase coherent transport through the device. It may well be a limitation in the performance of such devices.

In fact, our 0.1 μm gate width device is quite nearly a device formed on a quantum wire, since the width is also comparable to the inelastic mean free path of the carriers. Thus, as the gate potential is varied, one may well expect to see (even at room temperature) conductance fluctuations arising from the effects discussed above, as well as other quantum interference effects in the device. These conductance fluctuations appear as noise, but are not temporal variations—they arise in d.c. measurements and are quite repeatable. Such fluctuations, and their effects on device performance and behavior, can only be modeled with full quantum mechanical transport, and electrostatic, models.

3. Open Systems and Contacts

The implications of (4) are that the conductance of a localized tunneling (or scattering) barrier can be calculated from its transmissive behavior between two reservoirs.[74] If there are reservoirs on the left and right-hand sides of the transmissive region, which may be considered for the present as the *device*, it may be assumed that the system is in steady-state thermodynamic equilibrium deep within the reservoirs. The device resistance is then composed of the active region and the contacts, which connect the latter to the reservoirs. In a quantum mechanical sense, this is represented by the incident and out-going wave functions, their occupation probabilities, as well as the multi-dimensional transport through the system, and the included inelastic processes. Nevertheless, it is still possible, in principle, to calculate the transmissivity between the incident wave functions and output wave functions in the exit reservoir. Even if the transmission through the active region is totally elastic (the ballistic transport of the next paragraph), dissipation and ultimately irreversibility occurs through relaxation in the exit reservoir and the contact region adjacent to it. While physics normally considers closed systems, it is the macroscopic open system with its contacts, and reservoirs that are important to the consideration of devices.[75] Indeed, simulations of electron wave packet transport through quantum wires are sensitive to the details of the treatment of the reservoirs,[76] a result that is also known for studies of weak localization and universal conductance fluctuations in mesoscopic devices.[77]

This becomes more important in devices, as the problems of devices intrinsically involve open systems. As illustrated in Fig. 2, carriers within the device are interacting with a reservoir at each end. The electrons or holes in these reservoirs have been characterized as either satisfying a Boltzmann or Fermi distribution, in equilibrium, or as a displaced distribution in a nonequilibrium state of bias. Such a characterization implies that a local equilibrium exists in the "contact" and is fraught with all of the uncertainties associated with this characterization. It will be quite necessary to discover just how the characterization of the reservoirs, or contacts, actually impacts the simulations of the devices. We will see in the simulations in the following sections, that the boundaries and contacts play a very essential role, a role that is well known in the semi-classical world.

A. Ballistic Transport

The idea of a quantum trajectory, which resembles the classical phase space trajectory, dates to the early ideas of de Broglie and his pilot waves. It was resurrected by Bohm[47] to explain considerable detail of his wave approach. Nevertheless, it has been difficult to incorporate these trajectory ideas within quantum mechanics, since the probabilistic interpretation of the wave function tends toward the lack of a well-defined single trajectory for the wave packet. This is reflected in the summation over probabilistically weighted trajectories in path integrals.[48] Nevertheless, there is a consistent interpretation of quantum mechanics using trajectories as its basis.[78] This becomes important when we want to talk about *ballistic* transport of carriers from one contact reservoir to another. If the distance between the two reservoirs is less than the elastic mean free path, then carriers *injected* into the active region from one reservoir will drift under the applied fields to the other reservoir. It is not convenient to think about this motion in any other manner than as the transport of the centroid of the carrier wave packet along a semi-classical trajectory. Indeed, ballistic transport theory may be set up by choosing the appropriate fields to accelerate the carriers through the device; the problem becomes completely non-trivial if the fields are computed in a fully self-consistent manner.[35] The principles behind the Landauer equation (4) are not dependent upon this view, as one may define the *channels* by various modes of a waveguide, but it is often convenient to think of the Landauer equation in this fashion. We will see this further below, in connection with transport in high magnetic fields. In general, the nature of ballistic transport goes beyond structures whose lengths are less than the elastic mean free path. Rather, the important length over which the transport is essentially ballistic is the *inelastic* mean free path, or phase breaking length. That is, the important length is that over which the transport remains coherent. This was demonstrated by Büttiker,[79] who showed that transport through two series-connected phase-coherent regions produced the normal additivity of resistances only if an inelastic scattering process occured between the two regions. In the absence of this inelastic process, coherent addition of the two regions resulted. Surprizingly, if the overall transmission probability were low, then the addition of weak inelastic processes actually lowered the overall resistance, but the classical additivity of resistances was recovered as the degree of inelastic scattering was increased. It is thought that this initial lowering results from resonant transmission through the scatter, and this resonance produced higher transmission, as this effect did not occur for transmission probabilities of the order of 0.5.

The study of ballistic transport in quantum waveguide structures, in which a coherent structure supporting only a few occupied channels (a channel is one mode of transverse quantization) was placed between two reservoirs, has been pursued by a number of authors. One reason for this, is that the modification of (4) for the case of multiple modes has been somewhat controversial. To be sure, the overall conductivity is related to the overall transmission matrix of the multi-mode structure.[80] Yet the number of probes (side-arms are often added as voltage contacts, while the reservoirs serve as current contacts) affects the resultant formulae in many cases, and the result of reversing the magnetic field in a multi-probe measurement must yield the proper symmetries, consistent with the Onsager relations.[81] Indeed, if we have a four-terminal structure, with terminals 1 and 2 being used to provide the source and sink of current, and terminals 3 and 4 being used to measure the voltage, then (4) can be generalized to

$$R_{12,34} = \frac{h}{2e^2} \frac{1}{T} \frac{T_{31}T_{42} - T_{32}T_{41}}{(T_{31} + T_{32})(T_{41} + T_{42})} \tag{5}$$

and $T = T_{12} = T_{21}$. Clearly, if the last fraction is ignored, then (4) is recovered.

There are caveats to these equations as well. The "ballistic leads," e.g. the regions between the reservoir/contacts and the active region of measurement, must be connected in a nonreflective manner with the reservoirs, and the electro-chemical potentials must be measured deep in the reservoirs to assure thermalization of the carriers. If this geometry is not respected, then deviations can occur in measurements, and even in some theories, and this is important to considerations of potentials.[82] Consequently, the study of the existence of the proper transmission formula and resonances in the transmission under certain conditions, continues to be a topic of some discussion.[83] In addition, considerable effort is now being expended on the study of the high frequency forms of the conductance of these ballistic structures,[84] as well as the effect that bends in the ballistic structure play in the overall conductance.[85]

One of the most interesting aspects of the Landauer equation is the resulting experimental observation of quantized conductance through a constriction which could be varied by the gate voltage. Indeed, steps in the conductance were found in exact agreement with (4), as the measurements were essentially a two-terminal measurement.[86,87] However, the most extensive study of the multi-terminal version (5) has been in the quantum Hall effect.[88] In the presence of a high magnetic field, scattering of the carriers is suppressed as the ballistic trajectories are folded into Landau orbits, in which the essentially one-dimensional transport along the orbit hinders the scattering process.[89] Only those trajectories which reflect from the lateral boundaries move from one contact to another; these edge-located carriers are essentially what are now called edge states.[90–92] In a great many studies, which are not the central point of this review, the multi-terminal version (5) has now been verified. It should also be pointed out that, although (4) and (5) are basically obtained for non-interacting electrons, the results are not significantly affected by the presence of carrier-carrier interactions.[93,94]

While nearly all of the above ballistic electron studies have been carried out at low temperature, the basic nature of ballistic transport carries through from the semi-classical regime, and supports the basic trajectory nature of the transport of carriers, even in the quantum regime. This is likely to be an important consequence for the small semiconductor devices, in which we envision a need to include detailed quantum transport models. The ballistic transport provides one limit of the transport process and must be reflected in accurate models.

B. Role of the Boundaries and Contacts

One of the important consequences from (5) is that the actual resistance, or conductance, that is measured is dependent upon the details of the probes that are connected to the conducting channel. That is, it specifically depends upon whether there are two or four probes, or the details about how well the probes absorb or reflect incoming ballistic trajectories. In general, there are a large number of possible measurement probe geometries, and each delivers different possible overall conductances.[95] In short, one way of looking at this problem is that there is a variability in the possible voltages measured, for a given current, and it is important to know just where in space the voltage measurements are made. A slightly different view of this is that any particular (and certainly small) device is actually embedded within its environment. The boundary effects play an essential role in determining the physical properties of semiconductor quantum wires.[96] The performance of the device is not usually separable from its environment, and the environment can in fact completely determine the performance of the device.[12] Specific studies, with regard e.g. to resonant tunneling diodes clearly show this is the case,[97] a point that we will continue to see in the simulations discussed in later chapters.

The role of the interaction between the contacts (or probes or reservoirs, as the case may be) and the device can be significant. Indeed, it is actually possible for scatterers near the contact to induce oscillations in the electrochemical potential,[98] which further complicates the contact potential drop. The importance of the contacts and probes is best exemplified in the case of the quantum Hall effect.[99,100] Since the edge states are the penultimate ballistic (and non-dissipative) channel, the entire conductance and voltage distribution depends upon the details of the current and potential probes. The nature of this interaction can in fact be studied by varying the confinement potential to study the transition from local (classical) to non-local, ballistic transport.[101]

The idea of environment must be extended beyond just the concepts of probes and contacts. Indeed, it is the entire environment of a particular device that can lead to changes of device behavior. The presence of continuous devices opens the door to transfer between such devices, which has been especially studied in coupled quantum wires.[102] Another important scattering process, in addition to internal scatterers within the device, is interaction with the interface modes of the lattice.[103] In fact, in many cases the scattering from remote and interface modes may be more likely than scattering within the active device region, simply because of the fraction of phase space sampled by any particle wave function inside the device may be smaller than that part outside. This is also true for the surface, which may have extensive influence on the nature of the wave packets.[104] We will examine the specific formulation that leads to the detailing of the device-environment interaction in a great deal more depth later.

4. Some Potentially Important Quantum Devices

It has already been pointed out above that the normal Si MOSFET actually incorporates quantization within the channel, in the direction normal to the interface between the oxide and the Si inversion layer. Such quantization also appears in the GaAs/AlGaAs high-electron-mobility transistor (HEMT), in a similar fashion. Carriers are created by dopants placed in the AlGaAs, and these carriers transfer into the GaAs to create an accumulation layer on the GaAs side of the heterojunction interface. In both cases, the inversion/accumulation layers are created in a self-consistent potential with the actual size (thickness) of the layer being larger than the classical width due to the wave function of the carriers. This leads to a number of observable quantum effects in these devices, but which occur mostly at low temperatures. Moreover, much of these effects are only of second order in the transport properties. Nevertheless, full understanding of the ultra-small device will require a more advanced quantum transport treatment. As was also mentioned above, one of the most obvious quantum effects that can occur in the transport for ultra-short gate lengths is tunneling through the gate depletion region, and study of this effect is impossible in a classical treatment.

Tunneling is a fully quantum mechanical process in which a carrier penetrates into and traverses a barrier region, where the amplitude of the barrier exceeds the kinetic energy of the carrier. It first became of interest in semiconductors in highly-doped $p-n$ junctions, where the conduction band on the n-type side lay below the valence band edge on the p-type side (so-called degenerately-doped junctions).[105] The theory of such interband tunneling, which can also occur in semiconductors under very high electric fields (where it is often referred to as Zener tunneling) has been worked out over many decades, and has been reviewed extensively.[106-108] Even with a long history of work, there remain questions about the details of real tunneling processes in the presence of dissipative mechanisms,[109,110] and the *tunneling time*, the physical time required for a carrier to move from one side of the tunneling barrier to the other, remains quite controversial.[111] Tunneling is also the basis of the scanning tunneling microscope,[112] a new method of studying surfaces with atomic resolution. Nevertheless, the key device of interest at present is the resonant-tunneling diode.

A. The Resonant-Tunneling Diode

The idea of using two tunneling barriers, within a single conduction (or valence) band, is a relatively old idea. However, the concept of using band-gap engineering with semiconductor heterojunctions to create realistic barriers is only a few decades old,[113] and such structures have clearly shown negative differential conductivity as soon as they were made.[114] The basic idea is shown in Fig. 3. The barriers are formed by thin layers of AlGaAs, and the well and boundary layers are formed from GaAs. A quasi-bound state forms in the well layer. With no applied bias the tunneling through the structure is greatly reduced, as the tunneling length must extend over the entire width of the three barrier and well layers. On the other hand, when an applied bias is present, the anode layer and well layer are pulled to lower energies. When the quantum well level becomes degenerate with the occupied conduction band states in the emitter layer (Fig. 3b), current begins to flow through the structure, since there is now a resonant state available to the electrons, and the tunneling distance is now just that of the first barrier. If the two barriers were equal when the quantum level aligns with the emitter filled states, the transmission coefficient would rise to unity. When further bias is applied, so that the quantum level in the well drops below the conduction band edge of the emitter, current no longer flows. Thus, current flows only for a finite range of bias, and negative resistance is obtained on the high voltage side of this region.[115] We return to this below, with a more detailed description in each of the following sections, and include a discussion of non-resonant tunneling as well.

More recently, the resonant-tunneling diode has found applications in microwave circuits for amplification and oscillation.[116] Throughout the development of the resonant-tunneling diode, there have been a number of controversies. The first was whether the electrons tunneled completely through the structure coherently (in one step) or sequentially (in two steps).[117] In the end, it turns out that the actual current seems to be independent of this, but the only certainly sequential processes involve scattering of the carriers, especially when the scattering is inelastic, as has been determined by a number of relatively simple calculations.[118-121] The second controversy, and perhaps more meaningful, was over the role of trapped charge in the quantum well. For increasing bias, the quantum well has to be emptied as the current shuts off.. When the bias is again reduced, the well must be charged to begin the current. This suggests that there should be some hysterysis in the current-voltage curve.[122] To really study this effect, one needs self-consistent calculations for the current-voltage curve. Many of these have been carried out, and will be discussed in the later sections of this paper. Nevertheless, some more straight-forward approaches have also appeared; more straight-forward only in the sense that they try not to get involved in the detailed calculations of quantum transport.[123,124]

The resonant-tunneling diode is the "fruit-fly" for quantum transport studies, since the classical description cannot provide any insight into the process—the tunneling process is the key ingredient. Thus, we will see it treated again and again in the discussions below. In addition to the current, charge within the well, and the overall device characteristics, interest has focused recently on simple evaluations of the high-frequency conductivity and the noise.[125,126] When the resonant-tunneling diode is also laterally confined,[127,128] one gets a quantum dot (the laterally confined quantum well) with quite a complicated level structure within the dot.

B. Quantum Dots

The creation of lateral patterning to create an isolated region in which electrons (or holes) can be localized has led to considerable effort in the study of quantum dots and the electronic structure in these dots.[51,129] Usually, as in solids, the energy structure is calculated for the one-electron states, but the consideration of the carrier spin[130,131] and multi-particle states has also occured.[132] The quantum dot is usually thought of a localized region defined by gate potentials, as shown in Fig. 4, but it can actually be a waveguide resonator attached to an electron (or hole) waveguide.[133] The quantum dot is an interesting mesoscopic structure in its own right, as it is a mini-Aharonov-Bohm ring when edge states cycle through the structure,[134] and it can occur, and be studied, in a variety of manners, not the least of which is by STM probing.[135] We now are beginning to see studies of various "interactions" in the dot to specifically study its properties; e.g., recently Feng *et al.*[136] have created an additional gate controlled region in the center of the dot so as to act as an "impurity."

Obviously, the central quantum well of a resonant-tunneling diode also plays the role of a quantum dot. In fact, many aspects of quantum dots coupled to waveguides, or other probing regions, play much the same role as the tunneling coupling. However, the quantum dots can also be coupled capacitively, which is a classical interaction. Quantum dot effects begin to occur when the capacitors coupling the central region to its environment begin to be sufficiently small that the change in energy of the capacitor, when one electron transits it, is larger than the thermal energy or any bias energy. This is the so-called Coulomb blockade regime, and leads to the field of single-electron tunneling (SET).[137] One advantage of the use of capacitively-coupled dots is the ability to modulate the barriers and produce various device-like effects through this gate modulation.[138–140] Others have reversed this to use the oscillating barriers to actually study tunneling through these barriers.[141,142] Noise has also been suggested as a mechanism to study the tunneling properties of the barriers themselves.[143]

For the purposes here, the quantum dot is another interesting variant of the resonant-tunneling diode, and we will see several approaches to treat the detailed transport through these devices. It should be pointed out, though, that arrays of quantum dots form an interesting lateral surface superlattice, which leads to a number of other interested new physical effects, particularly in the magnetotransport.[144] Random arrays of the SET devices are a major candidate for future logic applications in the ultra-small regime.[137,145] It should also be pointed out that multiple quantum dots illustrate SET effects even in Si MOS structures,[146] a result that is expected from semi-classical modeling.[147]

II. The Quantum Equations

Although there are different formulations of quantum mechanics, nearly all approaches which lead to modeling of semiconductor devices derive from the Schrödinger equation

$$i\hbar\frac{\partial\Psi}{\partial t} = -\frac{\hbar^2}{2m}\frac{\partial^2\Psi}{\partial x^2} + V(x)\Psi \tag{6}$$

in one dimension (here taken to be x) so that $\Psi = \Psi(x, t)$. It should be noted that we have taken a particular form for the momentum, in that it is assumed that the particle energy is quadratic in the momentum. In devices, modeling usually proceeds from another formulation, which arises from (6), as will be shown below. The form of (6) is dissipationless, since the potential is normally the applied or built-in electrostatic potential. Although other forms are usually used, there has been work to actually apply the Schrödinger directly in simulations. For this, it must usually be assumed that the length of the region being simulated is considerably smaller than any characteristic dissipation length. Various characteristic lengths are important in the quantum mechanical description of transport.[148] Of most interest are the elastic mean free path, which describes a characteristic scattering length for elastic processes which do not break the phase coherence or relax the energy, and the inelastic mean free path, which describes processes which do break the phase coherence. There can be processes which break phase coherence, but do not relax the energy, although almost all energy relaxing processes break the phase coherence. Processes which can break phase coherence without relaxing the energy can arise from elastic processes that are sufficiently strong that they introduce localization. Thouless[149] suggested that one should relate the inelastic mean free path to the inelastic mean free time as

$$L_\phi = \sqrt{D\tau_{in}} \,, \tag{7}$$

where D is the carrier diffusion constant (it is assumed that the transport is diffusive, which implies that it is not ballistic or that there is considerable elastic scattering occuring within this length). Generally, it is still true that there is not a particularly good theoretical basis for calculating L_ϕ as yet.[148] Indeed, most estimates for its value are taken from experimental studies of the material in a particular device configuration.[15]

The basic concepts of transport in mesoscopic systems in the presence of localized scatterers can be traced to Landauer.[55] It is now recognized that slowly varying elastic potentials can lead to localization, and hence phase breaking in the system.[150,151] More importantly, there is a wealth of work now that clearly shows that the onset of inelastic scattering will suppress many of the quantum effects that are of interest; e.g. quantum interference effects.[152] There are many techniques to now simulate this, even with the Schrödinger equation.[153,154] Indeed, the role of scattering by large energy exchange processes, such as optical phonons, has clearly been demonstrated in studies of the DBRTD.[118-120,155-157]

The treatment of transport with the Schrödinger equation has followed several approaches. In one case, the scattering matrix formulation utilized by Büttiker[84] has been used to study simple waveguides in which elastic scatterers have been imbedded.[158] Here, a new concept has been introduced, and that is that the wave nature of the electron can be used to treat the transport of the electron as a guided wave problem,[74] just as in the case of microwave waveguides. The approach here uses (4) with the total transmission defined as

$$T = \sum_{n,m} T_{nm} \, , \tag{8}$$

where T_{nm} is the transmission from mode m of the input to mode n of the output contact. To develop this, it is usually assumed that the waveguide is created in an otherwise quasi-two-dimensional electron gas. The experimental waveguide itself can be defined either by physically creating a waveguide region by reactive-ion etching or by defining it electrostatically with lateral gates.[13] Then it is possible to write Schrödinger's equation in two dimensions as (time independent, however)

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\Psi(x,y)+V(x,y)\Psi(x,y)=E\Psi(x,y) \, , \tag{9}$$

with

$$V(x,y) = V_c(y) + V_{appl}(x,y) \, , \tag{10}$$

and the first term on the right-hand side is the confinement potential defining the lateral extent of the waveguide while the last term is any applied potential describing bias or impurities, etc. The general solution of the wave function in any small region (these regions are then connected together[159]) over which the lateral confinement potential is constant (which means that the waveguide has uniform properties) is given by

$$\Psi(x,y) = \sum_n \phi_n(x)\chi_n(y) \, , \tag{11}$$

where, in general,

$$\chi_n(y) = \sqrt{\frac{2}{W}}\sin\left(\frac{n\pi y}{W}\right) \tag{12}$$

for hard-wall confinement of the waveguide (in hard wall cases, it is usually assumed the wave function vanishes at the confinement wall). Other approaches sometimes use soft walls with quadratic potentials in which the lateral modes are described by harmonic oscillator wave functions.

The longitudinal modes are described, in general, by a combination of forward and backward propagating plane waves, as

$$\phi_n(x) = a_n e^{\gamma_n x} + b_n e^{-\gamma_n x} \, , \tag{13}$$

where γ_n is the propagation constant. If the mode is a propagating mode, then $\gamma_n = ik_n$ and describes the wave nature of the mode. If, on the other hand, the mode is evanescent, then γ_n is a real quantity describing the decay of the mode. ***It is very important to note that proper inclusion of the evanescent modes is very important in studying waveguide discontinuities by this method, just as it is in microwave waveguides.*** At the interface between two regions, in each of which the mode properties are uniform, the total wave function and its derivative are matched across the interface (note that if there is an applied potential at the interface site, the normal derivative is discontinuous by an amount determined by this potential[160]). This approach has been used to study the role of rounded corners at crossing waveguides,[161] waveguide stubs,[162] and the effect of impurities in the stub region,[163,164] as well as other configurations mentioned below.

One particularly interesting application is the study of a waveguide with a double constriction. That is, two narrow waveguides are separated by a wide region, and contacted with wider reservoirs on the ends, as shown in Fig. 5(a).[165,166] This structure is the waveguide equivalent of the DBRTD described previously. Modes are allowed in the wide central region which are below cutoff in the constricted regions, and this allows for tunneling into the central region, if the constrictions are sufficiently short, and consequently a negative-differential conductance can be obtained in the structure. In Fig. 5(b), the overall transmission probability is shown for such a structure as the energy of the incident wave is varied. This may be used in a conventional tunneling calculation to determine the current-voltage characteristics, and the resonance peak in the figure allows for the existence of the negative-differential conductance in these characteristics.[165]

It should be noted that the waveguide mode matching technique is fully compatible with formulation of scattering matrices,[167] although one normally thinks of the matching of the wave function and its derivatives in terms of a transfer matrix approach. The waveguide approach has great versatility, so long as the active regions are easily defined in terms of waveguide sections. It has been applied to tunneling between different waveguides,[168,169] to the onset of localization arising from rough waveguide boundaries,[170] to bends,[171] and to resonances from multiple bends in the waveguide.[172] In the latter, results comparable to lattice Green's function approaches[85] have been obtained. One problem is that self-consistent solutions to the waveguide mode propagation have not been obtained, although the role of carrier-carrier scattering between various sub-band modes has been studied,[173,174] and the general many-body problem has been formulated.[175] Also, a generalized density-functional theory has been proposed that could be used to incorporate many-body effects in the waveguide theory.[176] In general, however, the modeling approaches discussed in the remainder of this review are better suited for incorporation of many-body effects and self-consistency with the Poisson equation for the potential.

5. The Density Matrix and Its Brethren

In general, one can solve (6) by assuming an expansion of the wave function in terms of a set of static basis functions which satisfy the time-independent equation as

$$H\psi_n = E_n\psi_n \tag{14}$$

in which E_n is the energy level corresponding to the particular basis function. Then, the total wave function can be written as

$$\Psi(\mathrm{r},t) = \sum_n c_n\psi_n(\mathrm{r})\exp\left(-\frac{iE_nt}{\hbar}\right) . \tag{15}$$

If we now multiply each side of this equation with $\psi_m^*(\mathrm{r})$ and integrate over the position, we can evaluate the coefficient in terms of the total wave function at any arbitrary time, which we here take to be t_0. Then, (15) can be rewritten as

$$\Psi(\mathrm{r},t) = \sum_n \int d\mathrm{r}'\psi_n^*(\mathrm{r}')\psi_n(\mathrm{r})\exp\left(-\frac{iE_n(t-t_0)}{\hbar}\right)\Psi(\mathrm{r}',t_0) . \tag{16}$$

This may be rewritten as

$$\Psi(\mathbf{r},t) = \int d\mathbf{r}' K(\mathbf{r},t;\mathbf{r}',t_0)\Psi(\mathbf{r}',t_0)\,. \tag{17}$$

Here, $K(\mathbf{r},t;\mathbf{r}',t_0)$ is our propagator kernal, or Green's function.

The kernal in (17) describes the general propagation of any initial wave function at time t_0 to any arbitray time t (which is normally $> t_0$, but not necessarily so). There are a number of methods of evaluating it, either by differential equations (which we pursue here), or by integral equations known as path integrals.[48,177] In general, the form shown here is for a system described fully by a well developed set of basis functions, which are characteristic of the entire problem. For example, it is often the case that the time is taken to be imaginary, in which the substitution $(t - t_0) \rightarrow -i\hbar\beta$, where $\beta = 1/k_BT$ is the inverse temperature, and the resulting form of the kernal is that of a system in thermal equilibrium.[178] In this case, we talk about (17) representing a simple mixture of pure states. The usual case is that the exponential is separated into the two temporal parts, and then each time-varying basis function is expanded in an arbitrary (but different) set of wave functions, so that we have a mixed system, and we write the kernal as

$$K(\mathbf{r},t;\mathbf{r}',t_0) = \sum_{n,m} c_{nm}\phi_m^*(\mathbf{r},t)\phi_n(\mathbf{r}',t_0)\,. \tag{18}$$

The equal-time version of this is termed the density matrix

$$\rho(\mathbf{r},\mathbf{r}',t) = \sum_{n,m} c_{nm}\phi_m^*(\mathbf{r},t)\phi_n(\mathbf{r}',t) = \Phi^*(\mathbf{r},t)\Phi(\mathbf{r}',t)\,. \tag{19}$$

There are many different (in detail) definitions of the density matrix. It can be defined just by the coefficients in the expansion, so that it is a c-number matrix.[179] It also appears as the thermal equilibrium form defined above (for the time-independent form)

$$\rho(\mathbf{r},\mathbf{r}') = \sum_{n} e^{-\beta E}\psi_n^*(\mathbf{r}')\psi_n(\mathbf{r})\,. \tag{20}$$

The last form of (19) defines it in terms of field operators, in which creation and annihilation operators replace the expansion coefficients, and these operators excite or de-excite each of the "modes" of the basis set. In any of these definitions, it is important to recall that the density matrix is the equal time version of the Green's function.

A. The Liouville and Bloch Equations

In general, the density matrix is best characterized (for the present argument) in terms of the field operator form, which is the last part of (19). The temporal equation of motion for the density matrix can then be developed using (6) as

$$i\hbar\frac{\partial\rho}{\partial t} = [H,\rho]\,, \tag{21}$$

or[180]

$$i\hbar\frac{\partial\rho}{\partial t} = \left[-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial \mathrm{r}} - \frac{\partial^2}{\partial \mathrm{r}'}\right) + V(\mathrm{r}) - V(\mathrm{r}')\right]\rho(\mathrm{r},\mathrm{r}',t)\,, \tag{22}$$

which is termed the Liouville equation. Sometimes, a higher order operator algebra is used, since the Hamiltonian H is an operator in the Hilbert space of the density matrix (defined by some basis set of functions). In this case, (21) can be written as

$$i\hbar\frac{\partial\rho}{\partial t} = \hat{H}\rho(\mathrm{r},\mathrm{r}',t)\,, \tag{23}$$

where $\hat{H}$ is a commutator-generating superoperator.[181,182] There is no problem in incorporating a dissipative term in the Hamiltonian, and treating it by perturbation theory. In fact, this is a quite viable method of treating irreversible transport, as has been discussed repeatedly.[183–185]

On the other hand, if we accept the general view of the density matrix represented by (20), then it is natural to introduce the time as an imaginary quantity $t \rightarrow -i\hbar\beta$, and

$$-\frac{\partial\rho}{\partial\beta} = \left[-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial \mathrm{r}} - \frac{\partial^2}{\partial \mathrm{r}'}\right) + V(\mathrm{r}) - V(\mathrm{r}')\right]\rho(\mathrm{r},\mathrm{r}')\,, \tag{24}$$

which is normally termed the Bloch equation in the symmetrized space of the density matrix. In a sense, this form is a quasi-steady state, or quasi-equilibrium, form in which the time variation is either non-existent or sufficiently slow as to not be important in the form of the statistical density matrix. There exist mathematical proofs that a unique monotonic solution of this equation exists for the density matrix.[186] Before passing on, it is also important to note that there exists an adjoint equation to (24), which arises from the anti-commutator relationship, as[178]

$$-\frac{\partial\rho}{\partial\beta} = \left[-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial \mathrm{r}} + \frac{\partial^2}{\partial \mathrm{r}'}\right) + V(\mathrm{r}) + V(\mathrm{r}')\right]\rho(\mathrm{r},\mathrm{r}')\,. \tag{25}$$

This will be used below in a discussion of the connection to semi-classical behavior. In many respects, it should be noted that the density matrix is quite often found to be self-adjoint. In Fig. 6, we show a plot of the density matrix for a DBRTD in the absence of bias. The main diagonal, for which $x = x'$ in this one-dimensional model, represents the density variation through the device. The off-diagonal parts represent the spatial correlation that exists in the system.

B. Wigner Functions and Green's Functions

The problem with the density matrix in many semiconductor problems is that it is defined only in real space, with the important quantum interference effects occuring between two separated points in space. Even so, it is a function of six variables, plus of course the time (or the temperature). In many cases, it would be convenient to describe things in terms of a phase space function, whose six variables arise from a single position vector and a momentum vector. While this is not the normal case in quantum mechanics, it certainly can be arranged.[40,187] To see how this is achieved, we rewrite (22) in terms of a new set of coordinates, the center-of-mass and difference coordinates, as

$$\mathbf{R} = \frac{1}{2}(\mathbf{r} + \mathbf{r}') \, , \; \mathbf{s} = (\mathbf{r} - \mathbf{r}') \, . \tag{26}$$

Then, the Liouville equation can be rewritten as

$$i\hbar\frac{\partial\rho}{\partial t} = \left[-\frac{\hbar^2}{m}\left(\frac{\partial^2}{\partial\mathbf{R}\partial\mathbf{s}}\right) + V(\mathbf{R}+\frac{\mathbf{s}}{2}) - V(\mathbf{R}-\frac{\mathbf{s}}{2})\right]\rho(\mathbf{R},\mathbf{s},t) \, . \tag{27}$$

If we now introduce the phase-space Wigner distribution,[40] in three spatial dimensions,

$$f_W(\mathbf{R},\mathbf{p},t) = \frac{1}{(i\hbar)^3}\int d^3\mathbf{p}\rho(\mathbf{R},s,t)e^{i\mathbf{p}\cdot\mathbf{s}/\hbar} \, , \tag{28}$$

which is often called the Weyl transform,[188-191] then the Liouville equation can be written as

$$\frac{\partial f_W}{\partial t} + \frac{1}{m}\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{R}}f_W - \frac{1}{i\hbar}[V(\mathbf{R} + \frac{i\hbar}{2}\frac{\partial}{\partial\mathbf{p}}) - V(\mathbf{R} + \frac{i\hbar}{2}\frac{\partial}{\partial\mathbf{p}})]f_W = 0 \, , \tag{29}$$

in the absence of any dissipative processes. This can be rewritten in a more useful form as

$$\frac{\partial f_W(\mathbf{R},\mathbf{p},t)}{\partial t}+\frac{1}{m}\mathbf{p}\cdot\frac{\partial}{\partial\mathbf{R}}f_W(\mathbf{R},\mathbf{p},t)-\frac{1}{\hbar^3}\int d^3\mathbf{P}\,W(\mathbf{R},\mathbf{P})f_W(\mathbf{R},\mathbf{p}+\mathbf{P},t)=0\ , \tag{30}$$

where

$$W(\mathbf{R},\mathbf{P})=\int d^3\mathbf{q}\sin\left(\frac{\mathbf{P}\cdot\mathbf{q}}{\hbar}\right)\left[V(\mathbf{R}+\frac{\mathbf{q}}{2})-V(\mathbf{R}-\frac{\mathbf{q}}{2})\right]\ . \tag{31}$$

The use of the Wigner function is particulary important in scattering problems,[192] and it clearly shows the transition to the semi-classical world.[193] Reviewing the above approach, it may be simply stated that the Wigner function is the Fourier transform, in the difference coordinate, of the density matrix. These are two equivalent methods of looking at a problem, one (the density matrix) is entirely in real space, while the second is in phase space and often has the classical behavior as a limit. In Fig. 7, the Wigner function for a DBRTD, in the absence of bias, is shown. This should be compared to the density matrix version in Fig. 6. Clearly, the Wigner function shows much of the behavior of Fig. 1, in which a non-classical behavior is encountered near the potential barriers. On the other hand, there is not a great deal of difference in the representations of Figs. 6 and 7.

In general, the Wigner function described by (28) is not positive definite. This is a consequence of the uncertainty relationship between position and momentum. If (28) is integrated over all momentum, then the square magnitude of the wave function results, and this is a positive definite quantity, being related to the density. By the same token, integrating (28) over all position provides the square magnitude of the momentum density, which is also a positive definite quantity. It has been proved that the Wigner function provides a smooth continuous solution to the equation of motion (30).[194] Because of its close relation to the semi-classical Boltzmann equation, it may also be shown that it provides a robust solution to the couple Liouville and Poisson equations,[195] and therefore is quite usable for device modeling. More recently, Arnold has shown that the solution of (30) remains a "physical Wigner function," in the sense that it is a mixed quantum state consisting of a combination of pure states with non-negative distribution weights, for all times $t > 0$.[196] In fact, however, the Wigner function shown in Fig. 7

is positive definite, and this is a general result for the equilibrium "ground state."[197]

In both the density matrix and the Wigner function, only a single time variable appears in the problem, as it was assumed that the two wave functions, or field operators, in (19) were to be evaluated at equal times. In essence, these two approaches build in correlations in space but do not consider that there may be correlations in the time domain. This approach does not have to be taken, and this leads to the concept of the use of Green's functions themselves to describe the behavior of quantum systems.[198] In general, one separates the kernal in the wave function's integral expression for the propagator into forward and reverse times in order to have different functions for retarded (forward in time) and advanced (backwards in time, in the simplest interpretation) behavior. We do this by introducing the retarded Green's function as (for fermions)

$$G_r(\mathbf{r},\mathbf{r}';t,t') = -i\Theta(t-t')\,\langle K(\mathbf{r},\mathbf{r}';t,t')\rangle = -i\Theta(t-t')\left\langle \Psi(\mathbf{r},t)\Psi^\dagger(\mathbf{r}',t')\right\rangle \ , \tag{32}$$

where the angle brackets have been added to symbolize an ensemble average, which is also the summation over the proper basis states. On the other hand, the advanced Green's function is given by

$$G_a(\mathbf{r},\mathbf{r}';t,t') = i\Theta(t'-t)\,\langle K(\mathbf{r},\mathbf{r}';t,t')\rangle = i\Theta(t'-t)\left\langle \Psi^\dagger(\mathbf{r}',t')\Psi(\mathbf{r},t)\right\rangle \ , \tag{33}$$

and one can then write the kernal itself as

$$\langle K(\mathbf{r},\mathbf{r}';t,t')\rangle = i\left[G_r(\mathbf{r},\mathbf{r}';t,t') - G_a(\mathbf{r},\mathbf{r}';t,t')\right] \ . \tag{34}$$

Finding the Green's functions from the Schrödinger equation, or from the Liouville equation, is not difficult for simple Hamiltonians, as for any quantum mechanical problem. Proceeding for complicated Hamiltonians, such as in the case of many-body interactions or electron-phonon interactions, is not so simple and a perturbation approach is usually used. However, this approach is not without its problems, in that the perturbation series is difficult to evaluate and may not converge. Generating the perturbation series usually relies upon the S-matrix expansion of the unitary operator[199]

$$\exp\left(-\frac{i}{\hbar}\int_{t'}^{t} d\eta V(\eta)\right) \ , \tag{35}$$

where $V(t)$ is the perturbing potential interaction in the interaction representation. In nearly all cases, it is necessary to expand any perturbation series in terms of wave functions, and Green's functions, in the absence of the perturbation, which means at $t \rightarrow -\infty$. In the equilibrium situation, we can take the opposite limit as well, for the upper limit of the integral in (35), as it is assumed that the system is in absolute equilibrium at the point $t \rightarrow \infty$ as well. In the non-equilibrium situation, which is the normal case in nearly all active semiconductor devices, the latter limit is just not allowed. Then one must seek a better approach, and this has been given by the real-time (non-equilibrium) Green's functions developed by Keldysh[200] and Kadanoff and Baym.[201]

To avoid the need to proceed to $t \rightarrow \infty$ in the perturbation series, a new time path for the real time functions was suggested by Blandin *et al.*[202] (There may well have been others, but this seems to be the work which put it in proper context.) This new contour is shown in Fig. 8, where the contour evolves from the equilibrium (thermal) Green's function at $t_0 - i\hbar\beta$ to a real-time function at t_0. The contour then extends in the forward time direction to $\max(t, t')$, hence returning in the anti-time ordered direction to t_0.[203] In many cases, one lets $t_0 \rightarrow -\infty$ if we are not interested in the initial transients of the system. The handling of the Green's function, when both wave functions are on either the upper or lower branch is straight-forward. On the other hand, when these two functions are on different branches, two new functions, the correlation functions must be defined.[204] These are the "less than" function

$$G^{<}(\mathbf{r}, \mathbf{r}'; t, t') = i\left\langle \Psi^{\dagger}(\mathbf{r}', t')\Psi(\mathbf{r}, t)\right\rangle \ , \tag{36}$$

which has the opposite sign for bosons, and the "greater than" function

$$G^{>}(\mathbf{r}, \mathbf{r}'; t, t') = -i\left\langle \Psi(\mathbf{r}, t)\Psi^{\dagger}(\mathbf{r}', t')\right\rangle \ . \tag{37}$$

In general, these four Green's functions are all that are needed to handle the complete nonequilibrium problem (in relatively lowest order, as will be discussed later), but it is often found that two other Green's functions are useful. These are the time-ordered and anti-time-ordered Green's functions, in which the ordering is in the positive time progression around the contour of Fig. 8. These two are

$$G_t(\mathbf{r}, \mathbf{r}'; t, t') = \Theta(t - t')G^{>}(\mathbf{r}, \mathbf{r}'; t, t') + \Theta(t' - t)G^{<}(\mathbf{r}, \mathbf{r}'; t, t') \tag{38}$$

and

$$G_{\bar{t}}(\mathbf{r},\mathbf{r}';t,t') = \Theta(t'-t)G^{>}(\mathbf{r},\mathbf{r}';t,t') + \Theta(t-t')G^{<}(\mathbf{r},\mathbf{r}';t,t') \,. \tag{39}$$

There are obviously relationships between these six Green's functions, and these can be expressed as

$$G_r = G_t - G^{<} = G^{>} - G_{\bar{t}} = \Theta(t-t')(G^{>} - G^{<}) \,, \tag{40}$$

$$G_a = G_t - G^{>} = G^{<} - G_{\bar{t}} = -\Theta(t'-t)(G^{>} - G^{<}) \,. \tag{41}$$

For systems that have been driven out of equilibrium, the ensemble average brackets, indicated in the definitions of the Green's functions, no longer signify thermodynamic averaging or averaging over the ground state (at $T = 0$), since the latter quantities are ill-defined. Instead, the bracket indicates that some average needs to be taken over the available states of the nonequilibrium system, in which these states are weighted by the nonequilibrium distribution.

The equation of motion for the Green's functions are basically derivable from the Liouville equation above. The development of this equation for the various Green's functions will be put off until Sec. V. Here, we note that there are many methods of collapsing the Green's functions into single-time functions, which lead to a variety of transport equations.[205] Let us consider how to arrive at the Wigner function from the Green's function. We note that the definition of the density matrix, that led to the Wigner function, is basically quite similar to that of $G^{<}$. This is the proper association, as the latter function relates to the nonequilibrium distribution function.[201,204] Thus, we introduce the center-of-mass and difference coordinates (26), and equivalent ones for time (with T the average time and τ the difference time). Then it is clear that

$$f_W(\mathbf{R},\mathbf{p},T) = \lim_{\tau\to 0} \int d^3\mathbf{s}\, G^{<}(\mathbf{R},\mathbf{s},T,\tau) e^{-i\mathbf{p}\cdot\mathbf{s}/\hbar} \,, \tag{42}$$

or

$$f_W(\mathbf{R},\mathbf{p},T) = \int d\omega\, G^{<}(\mathbf{R},\mathbf{p},T,\omega) \,, \tag{43}$$

where the difference time has been Fourier transformed into a frequency in the last expression. It is clear from this last expression that the difference coordinates, introduced into the Green's function as in the density matrix, are Fourier transformed, and the Wigner function obtained from the Green's function by averaging out the frequency (or energy) dependence. Thus, the crucial kinetic variable in the Wigner function is the momentum, and not the energy, although the two are certainly related through a dispersion relation.

C. Reduced Density Matrices and Projection Operators

In each of the descriptions that has been introduced here, the density matrix, the Wigner distribution, and the real-time Green's functions, it has more or less been assumed that one is dealing only with the electron system. Indeed, it has mainly been assumed that one is dealing with single noninteracting electrons so that an equation for the equivalent one-electron distribution function is adequate. Before proceeding, it is of interest to consider how a quasi-kinetic picture can be obtained for the equivalent one-electron density matrix (or other approach) from a general system in which the electrons and the lattice all contribute to the density matrix, which can be a many-electron (and many atom) function. In general, the system is described by the Hamiltonian

$$H = H_0 + H_F + H_L + H_{eL} \,, \tag{44}$$

where the terms on the right-hand side represent the electronic motion, the external fields (in the scalar potential gauge), the lattice motion, and the electron-phonon interaction, respectively. The latter can include the coulomb interaction between impurity atoms and the electrons. The Hamiltonian H_0 includes all of the appropriate many-body terms and energy shifts appropriate to the full electron many-body problem, the details of which will not be treated here. The field term represents the driving fields through a simple form $H_F = -e\mathbf{F} \cdot \mathbf{r}$.

The total density matrix ρ is defined over the entire system: electrons, lattice, and interaction.[179] If it is decided to represent ρ in terms of a complete set of eigenstates for the electron and lattice systems separately, the general wave function will be a product of the individual wave function basis sets. The total trace operation, which appears as the representation of the ensemble average for a matrix form of the density operator, can be separated into a succession of of separate trace operations Tr_L and Tr_e, which represent the partial traces, or partial ensemble averages, over the lattice and electron components, respectively. This allows us to define the electronic density matrix $\rho_e = Tr_L(\rho)$. It is probably worth noting at this point that it is not at all obvious that this decomposition of the density matrix into clearly definable electron and lattice contributions will hold except in the steady-state case.[179] In essence, the objections are equivalent to those that limit the use of the effective mass approximation to relatively long time scales for the interaction. We will explore this more in later sections.

The approximations above are immediately invoked by introducing (44) into the Liouville equation and then summing over the lattice degrees of freedom as described. This gives

$$i\hbar\frac{\partial\rho_e}{\partial t} = [H_0 + H_F, \rho_e] + Tr_L\{[H_{eL}, \rho]\} \,. \tag{45}$$

Clearly, the first term on the right-hand side is the electronic motion within whatever effective mass approximation may be suitable. The second-term is the electron-lattice interaction. The trace over the lattice coordinates is equivalent to the summation over the phonon wavevectors in the electon-phonon interaction.

At this point, it is important to project out the desired part of the electron density matrix, which is usually the one-electron equivalent density matrix. This will be done by the use of projection operators.[181] To begin, (45) is Laplace transformed, with z taken to be the Laplace transform variable conjugate to the time. Then

$$\left(s + \frac{i}{\hbar}\hat{H}_e\right)\rho_e = -\frac{i}{\hbar}Tr_L\{\hat{H}_{eL}\rho\} + \rho_e(0) \,, \tag{46}$$

where $H_e = H_0 + H_F$, and the superoperator notation for the commutators has been used. We now obtain the one-electron density matrix through the projection operator P through[182,185]

$$\rho_{e1} = P\rho_e \,,\; P^2 = P \,,\; Q = 1 - P \,. \tag{47}$$

This particular projection operator commutes with the trace over the lattice variables, so that we may define a scattering operator as

$$\hat{\Sigma}\rho_{e1} \sim Tr_L\{\hat{P}\hat{H}_{eL}\rho\} \,. \tag{48}$$

With these definitions, (46) can now be written as[12]

$$\rho_{e1}(s) = \hat{P}\frac{1}{i\hbar s - \hat{H}_e - \hat{\Sigma}}i\hbar\rho_e(0) \,. \tag{49}$$

It is clear at this point that one only needs products of various projections of the resolvent operator (the fraction term above), and the projection operator identity[30]

$$\frac{1}{s-\hat{H}}=\left(\hat{P}+\hat{Q}\frac{1}{s-\hat{Q}\hat{H}\hat{Q}}\hat{Q}\hat{H}\hat{P}\right)\frac{1}{s-\hat{C}-\hat{P}\hat{H}\hat{P}}\left(\hat{P}+\hat{P}\hat{H}\hat{Q}\frac{1}{s-\hat{Q}\hat{H}\hat{Q}}\hat{Q}\right)$$

$$+\ \hat{Q}\frac{1}{s-\hat{Q}\hat{H}\hat{Q}}\hat{Q}\,, \tag{50}$$

where

$$\hat{C}=\hat{P}\hat{H}\hat{Q}\frac{1}{s-\hat{Q}\hat{H}\hat{Q}}\hat{Q}\hat{H}\hat{P}\,. \tag{51}$$

The latter term connects a "diagonal" element to an off-diagonal element, and then reconnects them by the conjugate operation, so clearly relates to a second-order interaction of the one-electron density matrix through the electron-lattice interaction. This leads us to rewrite (48) as

$$\hat{\Sigma}\rho_{el}(s)=Tr_L\{\hat{P}(\hat{H}_{eL}+\hat{H}_e)\hat{Q}\frac{1}{i\hbar s-\hat{Q}(\hat{H}_{eL}+\hat{H}_e)\hat{Q}}\hat{Q}(\hat{H}_{eL}+\hat{H}_e)\hat{P}\rho\}\,. \tag{52}$$

The temporal equation is obtained again by retransforming the density matrix. This results in the quantum kinetic equation

$$\frac{\partial\rho_{el}}{\partial t}=-\frac{i}{\hbar}\hat{P}\hat{H}_e\hat{P}\rho_{el}(t)-\frac{1}{\hbar^2}\int_0^t\hat{\Sigma}(t-t')\rho_{el}(t')dt'\,. \tag{53}$$

Clearly, the first-term on the right-hand side gives rise to the spatial variations of the one-electron density matrix that appear in (22), including the potential terms. The last term on the right-hand side is the scattering term, which has been ignored in the discussions above. This term incorporates retardation of the scattering, which by itself is a great difference from the Boltzmann equaiton. On the other hand, if the scattering term $\hat{\Sigma}$ varies slowly, then the convolution integral can be separated, and the Boltzmann transport equation reasserts itself as the semi-classical limit. Indeed, to lowest order, the scattering integral, the last term on the right-hand side of (53) is readily shown to give the Fermi golden rule for perturbation theory. The essence of the argument here, is that the electron-phonon interaction, the electron-impurity interaction, and even the electron-electron interaction can all be treated by perturbation theory, so that the one-electron (or one-hole) treatments developed in the preceeding paragraphs are all the equilibrium, or zero-order, formulations to begin with, and the higher-order interactions will work to perturb these solutions.

6. The Kubo Formula and Langevin Equations

Although a kinetic theory, such as that of the above paragraphs, is usually used as the basis of transport theory, there are alternative approaches tied rather directly to the use of correlation functions, either within a linear response formalism or with a Langevin equation formalism. The latter has existed for quite some time, but the formal development of the linear response formalism is relatively recent, following the work of Kubo.[206] As this approach is often used, we will review it in this section and then relate it to the various moment equations that can be obtained from the kinetic transport equations.

A. The Kubo Formula and Correlation Functions

In the Kubo approach, it is desired to find the response of the coupled electron-phonon system to a time-dependent perturbation by calculating to lowest order the change in the density matrix. As earlier, the Hamiltonian is written in the form (44). The difference here is that it is the electric field term H_F that is taken to be the perturbing potential. The quantity that is of interest is the current response to this electric field; e.g., the field is the forcing function and the current is the response to that force. To show this in detail, the field (assumed to be an a.c. field for the moment) is written in the vector potential gauge, so that the perturbing Hamiltonian term may be written as

$$H_F = \int d^3\mathrm{r}\, \mathrm{A}(\mathrm{r},t)\cdot \mathrm{j}(\mathrm{r})\,, \tag{54}$$

where $\mathrm{j}(\mathrm{r})$ is the paramagnetic part of the symmetrized total current operator. Now, one can either transform everything into the interaction representation, keeping just the lowest order terms in the exponential expansion (35) for linear response, or expand the density matrix in the Liouville equation

(21). The latter approach is used, although the end result is the same. The density matrix is expanded as

$$\rho = \rho_0 + \delta\rho(t) \,, \tag{55}$$

where ρ_0 describes the system prior to the application of the perturbation and is thus a system in which the electrons and the lattice are in equilibrium with each other, and with any internal electric fields that may result from inhomogeneous distributions of dopants within the semiconductor device. The linearized Liouville equation is then given by

$$i\hbar\frac{\partial\delta\rho}{\partial t} = [H, \delta\rho(t)] - [\rho_0, H_F] \,, \tag{56}$$

$$i\hbar\frac{\partial\rho_0}{\partial t} = [H, \rho_0] = 0 \,. \tag{57}$$

Here, $H = H_0 + H_L + H_{eL}$. The linear perturbation in the density matrix can then be obtained to be

$$\delta\rho(t) = \frac{i}{\hbar}e^{-iHt/\hbar}\int_0^t dt'[\rho_0, H_F(t')]e^{iHt/\hbar} \,. \tag{58}$$

In this last equation, it has been assumed that the perturbing Hamiltonian has its own time variation (which is the case for the field in the vector potential gauge), but vanishes for negative times. The exponentials in the unperturbed Hamiltonian and the time variation of the perturbation will result in the current operators winding up as the interaction representation, although we did not begin with it. It should be noted that the equilibrium density matrix produces no current, so that

$$< \mathrm{J}(\mathrm{r}, t) > = Tr\{\mathrm{J}(\mathrm{r}, t)\delta\rho(t)\} = \frac{i}{\hbar}\int_0^t dt' Tr\{\rho_0[H_F(t'), \mathrm{j}(\mathrm{r}, t')]\} \,, \tag{59}$$

where the cyclic properties of the trace have been used, the displacement current has been ignored, and

$$\mathrm{j}(\mathrm{r}, t') = e^{-iHt/\hbar}\mathrm{j}(\mathrm{r})e^{iHt/\hbar}. \tag{60}$$

It is possible to uncouple the retardation that appears in (59), by taking the Fourier transform of this equation and recognizing that the electric field is related to the vector potential by $\mathbf{F}(\mathbf{r},\omega) = i\omega\mathbf{A}(\mathbf{r},\omega)$. This leads, for a homogeneous electric field, to the result for the conductivity

$$\sigma_{\alpha\beta}(\mathbf{r},\omega) = \frac{1}{\hbar\omega}\int d^3\mathbf{r}' Tr\{\rho_0[j_\alpha(\mathbf{r},\omega), j_\beta(\mathbf{r}',\omega)]\} \ , \tag{61}$$

which is homogeneous only for size scales large compared to the inelastic mean free path.[207] The term in the ensemble average (the trace operation) is a retarded two-particle Green's function. It has an imaginary part which must be canceled by the displacement current at equilibrium, since the conductance is a real quantity at zero frequency. There have been many discussions through the years about the accuracy and applicability of the Kubo formula (61),[208] but it has now been fairly well established that it is accurate and gives results that agree with other treatments. Kubo himself used the approach to give one of the first detailed quantum treatments of high magnetic field galvanomagnetic effects.[209]

One important consequence of the discussion here is that the current response, and the conductivity, is now defined in terms of the fluctuations of the current itself. In essence, this is just the correlation function (or Green's function) that appears in the integral, and is a verification of the importance of the fluctuation-dissipation theorem. The current that flows (the conductivity) is determined directly by the dissipation arising from the presence of the scattering processes and the electric field. The correlation function describes these fluctuations, and the Kubo formula is no more than a direct statement of this important theorem. In fact, it is a powerful technique with which to calculate the noise properties of mesoscopic (as well as macroscopic) conductors.[210] Since the correlation function involves the fluctuations due to scattering processes, it directly incorporates these interactions within the calculation of the two-particle Green's function. Indeed, the role of carrier-carrier scattering in weak localization,[211] and of impurity scattering in universal conductance fluctuations utilize this approach,[80,69] even in inhomogeneous field situations such as these. Even in the so-called ballistic limit, (61) can be extended to study the a.c. response of quantum wires effectively.[212]

The Kubo formula (61) is valid at finite temperatures and for transverse as well as longitudinal fields. At low frequencies, however, there is a problem related to the coefficients in front of the integral. This is related to the fact that the vector potential diverges at zero frequency is approached. To get around this problem, the derivation can be modified by the use of an approach outlined by Mahan.[213] For this approach, the commutator that appears above is rewritten using the identity

$$[F(-t),\rho_0] = i\hbar\rho_0 \int_0^{\beta} d\beta' \frac{\partial F(-t - i\hbar\beta')}{\partial t} , \tag{62}$$

which is easily proven using the properties of thermal Green's functions, with $\beta = 1/k_B T$. With this identity, the conductance (61) can be rewritten as

$$\sigma_{\alpha\beta}(\mathbf{r},0) = \int_0^{t} dt' \int_0^{\beta} d\beta' \int d^3\mathbf{r}' Tr\{\rho_0 j_\alpha(\mathbf{r}', -t - i\hbar\beta') j_\beta(\mathbf{r})\} . \tag{63}$$

Although this result no longer contains the frequency, it still contains a nonlocal integration over the position variables, and still assumes that the electric field is uniform throughout the active region. It should also be noted that the expectation value is now a proper correlation function and no longer contains a commutator product.

At this point, it is important to point out that we have made a significant shift in theoretical emphasis. Until this section, the entire set of derivations and considerations has focused upon the *streaming* terms of the kinetic equations, and not upon the *relaxation* and/or scattering terms. With the present discussion of the Kubo formula, the opposite has now occured, in that here we are focusing upon the relaxation/scattering terms that give rise to the fluctuations and the streaming terms have been buried in reaching the Kubo formula (63). To be sure, the two approaches are not separable, and one must still evaluate the two-particle correlation functions that are the heart of the Kubo formula, and details of the relaxation effects of these two-particle Green's functions will expose the overall response of the system.[214] This concentration upon the dissipative processes, rather than the details of the streaming terms will also be present in the next paragraph.

B. Retarded Langevin Equations

Another alternative to transport theory depends upon the (retarded) Langevin equation and the Onsager relations. In this approach, the time-rate of change of a dynamic variable, such as the velocity, is related to a dissipative function, or to a series of forcing terms through dissipative coefficients.[215] The general form of these equations is

$$\frac{d\alpha_i(t)}{dt} = \sum_k \int_0^t dt' K_{ik}(t') F_k[\alpha(t-t')] \,, \tag{64}$$

where $\alpha = \{\alpha_i\}$ is the set of all observables, suitably ensemble averaged to produce the c-number observable, and the F_k are the thermodynamic forces.[216] The memory functions K_{ik} are time-dependent correlation functions, just as in the Kubo formula. Indeed, in some sense, (64) is a generalized form of the Kubo formula. The general approach to finding the equations that make up the set (64) is to expand the density matrix in linear response, as done in (56) and then use this to compute ensemble averages of the dynamic variables, which is just the procedure used for the Kubo formula. It may be noted that the ordinary Langevin equation (absent the convolution integral in time) may be obtained if the thermodynamic forces vary slowly on the scale of the relaxation forces, and the memory functions vary sufficiently rapidly (and decay sufficiently rapidly) that the integral over time is not sensitive to the final time; e.g., the time integrals must be convergent.

Advantages to using the Langevin equation approach is the equal treatment of the frictional (dissipative) forces (which give a non-vanishing result in the memory function integration) and the fluctuating forces, which average to zero in any time integration.[217] Moreover, the generalized Langevin equation is a reduced description of the system, which can profit from the use of many approaches developed for statistical physics. While it can be argued that this approach does not now need to worry about the density matrix, the latter is inherent in the treatment, and one still must worry about the details of the evaluation of the memory functions, which in keeping with the Kubo formular are two-particle correlation functions. Nevertheless, it is possible to take the generalized Langevin equation beyond linear response to treat, for example, localization of the carriers[218] and high-electric-field transport.[32,219–222] Because one must inherently calculate the memory function as a many-particle interaction, it is easy to extend it to a many-body interaction.[223,224]

Over the past few years, an alternative approach has cropped up to generate these Langevin equations. This approach is given many names, but the most useful is *force balance*. We can understand this approach from the simple fact that, in analogy to the Liouville equation, the time variation of the expectation value of an operator, which itself is not an explicit function of time, is given by

$$\frac{\partial < \alpha >}{\partial t} = \frac{i}{\hbar} \langle [H, \alpha] \rangle \ . \tag{65}$$

At first sight, these results appear to be the same as those of the retarded Langevin equation above, and in some cases this is the proper interpretation. However, the manner in which these force balance equations has been used is to calculate the *resistivity*, and not the *conductivity*. Recall that it is the field (or the potentials) that are applied to the system, and it is the current which is the response of the system, so that the proper Kubo formula, or Langevin equation, deals with the conductivity or the momentum, respectively, for the response. Much of the literature on the force balance arises from Ting and his colleagues,[225] although Peters and Devreese[226] have also used this approach to compare the force balance with the Feynman path integral approach. In principle, the evaluation of the resistivity is conceptually thought to be easier since, for weak scattering, one only has to keep the low order terms. However, this approach has been criticized rather heavily, and there is an indication that the correct evaluation of the resistivity (and corresponding coefficients for other equations than the current) still requires an infinite summation over terms, even for weak scattering.[227] While the former authors feel that they have answered (successfully) such criticism, the approach remains controversial in many aspects, particularly as many approximations made in the earlier formulations reduced the scattering integrals to the Fermi golden rule and to Boltzmann transport.[228]

One problem with this approach is that (65) requires some sort of ensemble average to be performed, and *the question is just which ensemble is to be used*. Obviously, it is not the equilibrium ensemble, since there is no current in this ensemble. Thus, the non-equilibrium distribution, or density matrix, must be computed as part of the problem, and this part of the problem is often ignored. In particular, the nature of the distribution can affect the quantum corrections to the streaming terms in the quantum kinetic equations. On the other hand, more recent calculations, explicitly using a quasi-equilibrium density matrix (as will be discussed in the next section), yield results that are in keeping with other approaches.[229,230] With the quasi-equilibrium density matrix approach, the results seem to be equivalent representations of the

Retarded Langevin Equation and the Kubo formula.

These approaches have been applied to a variety of problems, in which estimates of the correlation functions have been made through explicit calculations, such as in the case of noise and high-electric-field transport.[231-235] Transport in quasi-two-dimensional systems and wires has also been considered,[236-239] as has nonlinear transport with nonequilibrium phonons.[240]

7. Boltzmann-Like Approaches

The Boltzmann transport equation (1) has been studied for a great many decades, and a considerable insight into its functioning now exists. On the other hand, quantum transport has been studied for a considerably smaller amount of time, and in most cases such insight is nonexistent. For this reason, many theoretical approaches have been aimed at trying to establish (1) as a result of a limiting process from a more fundamental quantum basis, that is, one attempts to determine a quantum Boltzmann equation.[241] In general, these approaches do not simply reduce to the normal Boltzmann transport equation, since the proper manner in which to take the limit has never been defined well. Nevertheless, these derivations have been pursued for quite some time.[242] In general, however, it is necessary to formulate the problem with one of the more exact quantum transport equations, and then connect one of the quantum distributions to the semi-classical Boltzmann distribution through some sort of ansatz.[204,243] This approach tends to work well when the major problem is one of high scattering rates in an otherwise weak scattering process.[244] In most cases, the results look just like the Boltzmann equation, although some novel approaches have been suggested.[245]

In most cases, the quantum transport may be cast into the form of the Boltzmann equation from e.g. real-time Green's functions. Then known forms for quantum animals such as the polarization[246] or polaron effect[247] can be used in the transport equation as scattering terms. In other cases, the major quantum mechanical effect is the dynamic change of the density of states, such as in Landau quantization, and this can be incorporated within (1) by using a multi-band picture,[248,249] although this is sometimes done through the force-balance equation, which is subject to the concerns mentioned above. Other effective approaches use the real-time Green's functions to treat the quantum transport, and the limit is approached through taking the equal time results, so that variations in densities are the main result.[250-252] In this latter case, the equation of interest is a moment of the kinetic transport equation, which relates it to the approach of the last section, although many higher order moments can be used as well.[250]

The major point to be made here is that the transition from quantum to classical (or semi-classical) transport must be approached carefully, and the proper route is not at all obvious. Consequently, some of the more obviously useful approaches are through moments of the transport equations themselves, as there is often a more obvious connection between the quantum and classical worlds. This was the basis of the previous sub-section, as well as the next. Nevertheless the transition often incorporates effects which are not present in the classical world and make the limiting process difficult to achieve.

8. Moment Equations—The Transition from Classical to Quantum

The discussion of the previous two sections, regarding how to achieve force balance (or other balance) equations in a method which does not utilize the rigor of the Kubo formula (or the equivalent retarded Langevin equation), suggests that one should approach this topic with a great deal of care. Foremost, it is important to know that the ensemble averages that must be computed in e.g. (64) depend upon the details of the ensemble, or distribution function, itself, so that this crucial computation must still be carried out. In many situations, it is possible to create a quasi-equilibrium distribution function, which is parameterized in the observables. For each of these parameters, an equation of motion, describing its temporal evolution, must be formulated—this leads to the set of so-called balance equations. This approach is quite old, having been basically studied for more than half a century (not counting the work on the classical Boltzmann transport equation). The structure of this is that the evolution must be decoupled from the initial condition, usually by inter-particle scattering, so that a quasi-equilibrium density matrix may be defined by its integral invariants.[253] Indeed, it may generally be said that the quasi-equilibrium density operator may be written as[254]

$$\rho = \exp\left[Z - \sum_k f_k P_k\right] , \tag{66}$$

where Z is the partition function, f_k is the integral invariant force, and P_k is the conjugate quantum mechanical operator. Examples of these are the Fermi energy, the average momentum, and the inverse temperature β for forces, which are proportional to the conjugate force for the number, momentum, and energy operators, respectively. To each of these pairs, we must have a "balance" equation which describes the temporal evolution of the appropriate quantity.[255] In many cases, these balance equations offer a more convenient method of solving for transport properties, or for modeling devices, than the more detailed full solutions of the appropriate quantum mechanical distribution.

Even with this approach, one is still faced with the manner in which the classical limit should be approached. Consider for example (27) which can be rewritten in the form (in the absence of scattering)

$$\frac{\partial \rho}{\partial t} - i\frac{\hbar}{m}\frac{\partial^2 \rho}{\partial \mathbf{R} \partial \mathbf{s}} + \frac{i}{\hbar}[V(\mathbf{R} + \frac{\mathbf{s}}{2}) - V(\mathbf{R} + \frac{\mathbf{s}}{2})]\rho(\mathbf{R}, \mathbf{s}) = 0 . \tag{67}$$

It would be nice to accept the classical limit of this equation as the Boltzmann equation (1). For this to be the case, however, several assumptions and limits must be introduced. One that is not too difficult to make is that the corresponding momentum is given by $\mathbf{p} = -i\hbar\frac{\partial}{\partial \mathbf{s}}$. Indeed, this is the result expected from the transformation into the Wigner formulation. The next term is more critical. To obtain the classical force term in (1) requires expanding the potential in a Taylor series, and then associating the difference coordinate with $\frac{i}{\hbar}\mathbf{s} \rightarrow -\frac{\partial}{\partial \mathbf{p}}$, which again is not unusual as it connects p and s as conjugate operators. The problem arises in the fact that *we must also limit the potential to be of no higher power than quadratic in the coordinates.* This is just not the normal case, and we must face the fact that if the potential varies with higher powers than quadratic, the quantum transport equation will not reduce to the classical one! There will be extra terms that represent these higher-order variations; indeed, these extra terms correspond to the Wigner-Kirkwood expansion used in estimating the Wigner transport equation. But this behavior is just the situation that is expected from the density variation shown in Fig. 1 above. It is clear that we do not approach the classical limit by simply letting $\hbar \to 0$; the limiting process is more involved than this. As a final illustration of this, we now note that the classical distribution function is finally obtained by $f = \lim_{s \to 0} \rho(\mathbf{R}, \mathbf{s})$, but this limit must be taken after all appropriate derivatives (with respect to the difference variables) are evaluated.

In many respects, the behavior of the quantum distribution, and the density, must incorporate a nonlocal *quantum potential*, which produces the behavior of Fig. 1. We introduced this topic in the early parts of this review, where we discussed the Wigner[40] and Bohm[47] potentials. As mentioned, other approaches have been based upon variational approaches,[49] or other estimations.[256] There has been a great deal of discussion in particular about the Bohm potential, as it tends to be linked closely to a different interpretation of quantum mechanics, which is somewhat controversial.[257] Nevertheless, it has demonstrated the ability to predict quantum energy levels and is useful in many applications,[258] particularly in mesoscopic devices.[259] Here, however, we will take a different approach to find an equivalent quantum potential. To this end, we write the adjoint of the Bloch equation in the center-of-mass and difference coordinates as (again, in the absence of scattering processes)

$$\frac{\partial \rho}{\partial \beta} = \left\{ \frac{\hbar^2}{8m} \frac{\partial^2}{\partial \mathbf{R}^2} + \frac{\hbar^2}{2m} \frac{\partial^2}{\partial \mathbf{s}^2} \right\} \rho - [\cosh(\frac{1}{2}\mathbf{s} \cdot \nabla) V(\mathbf{R})] \rho \,, \tag{68}$$

where

$$2[\cosh(\frac{1}{2}\mathbf{s} \cdot \nabla) V(\mathbf{R})] = [V(\mathbf{R} + \frac{\mathbf{s}}{2}) + V(\mathbf{R} - \frac{\mathbf{s}}{2})] \,. \tag{69}$$

Within the approximation that the logarithm of the density matrix may be approximated as linear in the potentials, it may be shown then that the density matrix may be found from (68) as[50]

$$\rho(\mathbf{R}, \mathbf{s}) \sim \exp\left\{ -\beta(W_Q + U_Q) - \frac{3}{2}\ln(\beta) + \frac{i\mathbf{p}_d \cdot \mathbf{s}}{\hbar} - \frac{m}{2\beta}\left(\frac{\mathbf{s}}{\hbar}\right)^2 \right\} \,, \tag{70}$$

where the first term is found from the equation

$$-\lambda_D^2 \frac{\partial^2 (W_Q + U_Q)}{\partial \mathbf{s}^2} + \mathbf{s} \cdot \frac{\partial (W_Q + U_Q)}{\partial \mathbf{s}} + W_Q + U_Q = -\frac{\hbar^2}{2m} \frac{\partial^2 \rho}{\partial \mathbf{s}^2} + W(\mathbf{R}, \mathbf{s}) \,, \tag{71}$$

and $W(\mathbf{R}, \mathbf{s})$ is given by (69). The solution to this is given as[50]

$$W_Q(\mathbf{R}, \mathbf{s}) + U_Q(\mathbf{R}, \mathbf{s}) = \int \frac{d^3 \mathbf{s}'}{4\pi \lambda_D^2 |\mathbf{s} - \mathbf{s}'|} \{ W(\mathbf{R}, \mathbf{s}') - \frac{\hbar^2}{2m} \frac{\partial^2 \rho}{\partial \mathbf{s}'^2} \} \exp[-\frac{(\mathbf{s} - \mathbf{s}')^2}{2\lambda_D^2}] \,. \tag{72}$$

Thus, the actual potential, and an additional density-dependent function, are smoothed by the exponential Green's function. Here, λ_D is the thermal de Broglie wavelength introduced earlier. The exact actual relation of (72) to either the Wigner or the Bohm potential is not currently known. However, it has now been shown that, in the limit of slowly varying potentials where only low order terms in the potential expansion are treated, (72) reduces to simply

$$\lim_{\mathbf{s}\to 0}[W_Q(\mathbf{R},\mathbf{s}) + U_Q(\mathbf{R},\mathbf{s})] \rightarrow V(\mathbf{R},0) + U_B(\mathbf{R},0) \,, \tag{73}$$

so that it is the Bohm potential that first modifies the statistical thermodynamics of the distribution function.

A. The Hydrodynamic Moment Equations

The density matrix has the usual characteristics that, in the limit $\mathbf{s} \rightarrow 0$, and with appropriate other limiting processes (e.g., in connection with $\hbar$), it becomes the normal density $n(\mathbf{R})$. The equation of motion for the density matrix, which should have some asymptotic connection with the Boltzmann equation for $n(\mathbf{R})$, is obtained from the Liouville equation, and is given from (67) as (in the absence of dissipation)

$$\frac{\partial\rho}{\partial t} - i\frac{\hbar}{m}\frac{\partial^2\rho}{\partial\mathbf{R}\partial\mathbf{s}} + \frac{2i}{\hbar}[\sinh(\frac{1}{2}\mathbf{s}\cdot\nabla)V(\mathbf{R})]\rho(\mathbf{R},\mathbf{s}) = 0 \,. \tag{74}$$

Here, the sine function has been obtained in the same manner as the cosine in the preceeding paragraph as a representation of the two displaced evaluations of the potential. We note that this last term leads only to odd orders of derivatives of the potential, as already noted, and the higher orders (higher than the first order) are clearly quantum correction terms as they do not appear in the Boltzmann equation. If we now take the limit of this equation as $\mathbf{s} \rightarrow 0$, we find

$$\frac{\partial n}{\partial t} = \lim_{\mathbf{s}\to 0}[\frac{i\hbar}{m}\frac{\partial^2\rho}{\partial\mathbf{R}\partial\mathbf{s}}] = -\frac{1}{m}\nabla\cdot(\mathbf{p}_d n) \,, \tag{75}$$

where we have used the operator definition of the momentum given above, asserting that an averaging process takes place as well. This is the well-known continuity equation. However, there are some problems of interpretation, since in the absence of dissipation the transport must be reversible. In a strict sense, the distribution is symmetric across the diagonal axis ($s = 0$), so that the average momentum and the time derivative both actually vanish. The equation strictly has meaning when dissipation in the system (along with the proper driving forces) create asymmetries in the density matrix and drive it out of equilibrium. This does not mean that we cannot create an asymmetric distribution representing ballistic transport through a dissipation-free region. However, to do so requires special considerations in the contacts (hence the need for treatment of an open system) to maintain this special distribution function.[75] Then, to consider (75) as having meaning separate from the special situations of the contacts is improper, as the contacts provide the *sources* and *sinks* of nonequilibrium carriers that make up the asymmetric distribution. The dissipative terms, which will be discussed below, then provide these sources and sinks on a local scale throughout the dissipative region where transport is being considered. Indeed, it is impossible to define a diffusion or drift current from (74) without the dissipative terms, as the currents due to these processes must result from a careful balance of driving forces and dissipative forces. Nevertheless, we can use (75) to *define* the appropriate moment of the distribution. Within this interpretation, and limitations,

$$\mathbf{p}_d n(\mathbf{R}) = Tr\{-i\hbar\frac{\partial\rho}{\partial\mathbf{s}}\}\,, \tag{76}$$

and the trace is a local evaluation which produces the classical (or semiclassical) density variations.

In a similar fashion, the first-order moment equation can be developed by taking the derivative of (74) with respect to the difference variable. Then, upon passing to the limit, we find

$$\frac{\partial}{\partial t}(\mathbf{p}_d n) = \frac{\hbar^2}{2m}\nabla\cdot\left(\lim_{\mathbf{s}\to 0}\frac{\partial^2\rho}{\partial\mathbf{s}^2}\right) - n(\mathbf{R})\nabla V\,. \tag{77}$$

The first term on the right-hand side is the normal divergence of the "momentum pressure" tensor in the classical limit. As is often done in the semiclassical case, we may approximate this by taking the pressure tensor as an isotropic scalar quantity which leads to

$$-\frac{\hbar^2}{2m}\lim_{\mathbf{s}\to 0}\frac{\partial^2\rho}{\partial\mathbf{s}^2} \to < E > n(\mathbf{R})\,. \tag{78}$$

Using our quantum potential derived above, this latter quantity becomes

$$< E >= U_{eff} + \frac{3}{2\beta} + \frac{\mathbf{p}_d^2}{2m} , \tag{79}$$

where

$$U_{eff} = \lambda_D^2 \frac{\partial^2 (W_Q + U_Q)}{\partial \mathbf{s}^2} = (W_Q + U_Q) - \{W(\mathbf{R}) + \frac{\hbar^2}{2m}\frac{\partial^2 \rho}{\partial \mathbf{s}^2}\} . \tag{80}$$

Thus, ***the effective quantum potential is the difference between the smoothed value and the unsmoothed value.*** This means that if the driving functions, the potential itself and the density variations that respond to this potential, are slowly varying functions on the scale of the thermal de Broglie wavelength, the effective quantum potential goes away and is not a factor. In this sense, we recover the classical forms for the balance equations in the limit that the spatial variations become slow on the scale of the thermal de Broglie wavelength. In some sense, the first corrections appear at a WKB level of variation, as may be expected. In fact, in the limit in which the potential is slowly varying so that only the low orders of the potential expansion need be maintained, it is found that the effective potential reduces to the Wigner potential. Thus, while the statistical mechanics are governed by the Bohm potential in this limit (discussed above), the transport behavior is governed by the Wigner potential.

To obtain the energy equation, we need to take the second derivative of (74) with respect to the difference coordinate, and then pass to the limit of vanishing **s**. For this, however, we need to evaluate the third derivative of the density matrix, with respect to this difference coordinate. In keeping with previous approximations, this leads to (again, in the scalar approximation with no dissipation in the system) the energy equation (with no dissipation)

$$\frac{\partial}{\partial t}[< E > n(\mathbf{R})] = \nabla\cdot \left[\frac{i\hbar^3}{6m^2}\lim_{\mathbf{s}\to 0}\frac{\partial^3 \rho}{\partial \mathbf{s}^3}\right] = -\frac{\mathbf{p}_d}{m}\cdot(\nabla V) n(\mathbf{R}) , \tag{81}$$

and the term in large square brackets may be evaluated as

$$\frac{i\hbar^3}{6m^2}\lim_{\mathbf{s}\to 0}\frac{\partial^3 \rho}{\partial \mathbf{s}^3} = \frac{\mathbf{p}_d}{m}\left(\frac{\mathbf{p}_d^2}{2m} + U_{eff} + \frac{3}{2\beta}\right) n(\mathbf{R}) . \tag{82}$$

This particular form of the energy arises from the particularly simple manner in which the scalar approximation has been used to achieve the diagonal result. A more careful evaluation of the full tensor approach will yield a slightly different result, and some of the numerical factors can change (for example, the last term in the parentheses often appears as 5/2 rather than 3/2).[260]

We note in these developments that the effective quantum potential has contributions from both the spatial variation of the potential *and* from the density variation that is a response to that potential variation. This differs from earlier treatments, in that early WKB expansions easily obtained additional terms in a Taylor series expansion of the potential, with the assertion that this connected to the logarithm of the density through a simple classical partition function. This produced an effective quantum potential that is one-third of the Wigner potential (2).[40,178,261] Actually, it is important to note that the form of the effective potential arises from the manner in which the second and third moments of the density matrix in the difference coordinate, as represented by these respective derivatives, are related to the potential and average momentum and energy.[46,262] Nevertheless, it is not possible to simply relate the higher-order derivatives of the potential directly to the density matrix, as the actual effective potential arises from the self-consistent interactions between the density and the potential.

B. Applications in Modeling Devices

The classical hydrodynamic equations have a long history in semiconductor device modeling.[263–266] The extension of these approaches to include the modifications of the statistical mechanics due to the quantum effects has allowed one to begin to model smaller devices without the need to use a fully quantum kinetic approach.[46,267,268] Here, we concentrate on the effect played by the effective quantum potential in the device. Grubin *et al.*[261,267] and Gardner[260] have used the quantum hydrodynamic equations, with the effective potential taking the value of 1/3 of the Wigner potential, to model a single tunneling barrier and a DBRTD. In particular, the former authors compared the results with exact calculations using the density matrix directly, and found good agreement between the two approaches. For the transport, a Fokker-Plank form of dissipation was added to (74) in the form

$$\left.\frac{\partial \rho}{\partial t}\right|_{coll} = -\frac{2s}{\tau}\frac{\partial \rho}{\partial s} - \frac{4Ds^2}{\hbar^2}\rho\,. \tag{83}$$

In Fig. 9, the effective potential is shown in the region of the single barriers described by

$$V_b(x) = \frac{V_0}{2}\left[\tanh\left(\frac{x+a}{b}\right) - \tanh\left(\frac{x-a}{b}\right)\right]\,, \tag{84}$$

where $V_0 = 0.5V$, $a = 15nm$, and $b = 1.315nm$. It can be seen from this figure that the effective potential has a peak that is almost 60% of the actual potential in a region on either side of the barrier.

Using the Wigner potential itself, Zhou and Ferry[269] have modeled MESFETs in GaAs and SiC, and HEMTs in the AlGaAs/GaAs and strained-Si/GeSi heterojunction systems for ultra-submicron gate lengths ($L_g < 0.1\mu m$). For these studies, the Wigner potential form for the effective potential was used, and a relaxation-time approximation was used for the dissipation, as

$$\left.\frac{\partial \rho}{\partial t}\right|_{coll} = -\frac{\rho(\mathbf{R}) - \rho_0(\mathbf{R})}{\tau}\,. \tag{85}$$

Separate relaxation times are used for momentum and energy, and the energy dependent values for these are found from ensemble Monte Carlo calculations for homogeneous material. It should be remarked that $\rho_0(\mathbf{R})$ is *not* the equilibrium Fermi-Dirac distribution, but is the actual spatially dependent diagonal density matrix found self-consistently with the built-in potentials from gates, doping variations, or barriers. For example, in the DBRTD, it is the density appropriate to either Fig. 6 for the density matrix or Fig. 7 for the Wigner distribution (which gives essentially the same hydrodynamic equations). Normally, for example, one expects (85) to vanish when the continuity equation (75) (the lowest-order moment equation) is computed. This is true, however, only so long as the local density remains that appropriate to the unbiased self-consistent potential. In active devices, the density can deviate from this latter value when applied potentials exist, and the scattering decay term in the continuity equation corresponds to the diffusive restoring forces that work to reduce the charge fluctuations. In essence, these are spatial generation-recombination terms that correspond to the movement of charge under the scattering induced diffusive processes. We return to this point later.

In Fig. 10, the effective quantum potential is shown for the GaAs MESFET. Here, the potential is large at the interface between the active layer and the semi-insulating substrate, as well as around the gate depletion layer. The quantum potential works to retard channel pinchoff, leading to a generally higher current through the device. In Fig. 11, the carrier density profile in the plane normal to the gate (and under the gate) is shown for a GaAlAs/GaAs HEMT. The quantization in the depleted GaAlAs layer leads to a reduction in the real-space transfer of carriers out of the active channel. On the other hand, the tendency of the quantum potential to retard pinchoff also works to reduce this real-space transfer. These two effects lead to a current enhancement of as much as 20% over the classical results (where the quantum potential is set to zero). It is clear that the quantum effects will be quite important in devices with gate lengths below 0.1 μm.[270] On the other hand, it is clear that we have just begun to examine the quantum effects that can occur in such small devices. All of the above simulations have been either one- or two-dimensional simulations. As discussed in the introduction, a small device of 0.05×0.1 μm gate (length times width) will have only of order 100 electrons in the active region. This suggests that the non-uniformities, and the individualness, of the impurities and electrons, will lead to significant quantum fluctuations.[271] In fact, Monte Carlo estimates of the inelastic mean free path in Si at room temperature suggests that this quantity is of the order of 0.1 μm at low fields and ≥ 0.05 μm for fields up to 10^5 V/cm.[272] This suggests that effects such as universal conductance fluctuations may be observed at room temperature in such small devices. To fully investigate these effects, three-dimensional simulations are required in which the charge is a set of spatially localized impurity atoms and the corresponding electrons, rather than smoothed distributions. Or, more exact quantum mechanical simulation approaches are required.

III. Modeling with the Density Matrix

In writing the density matrix as in (19) or (20), it is assumed to be a single particle quantity. In fact, the density matrix starts life as a many-body quantity,[273] and it must be projected to a description as a single-particle density matrix. This process is the same as followed in Sec. II.5.c. In essence, the projection operator that accomplishes this task integrates out the variables of all particles except for one, leaving a two-particle interaction which couples the one-particle density matrix to a two-particle density matrix.[274] Approximations to the latter term particle-particle lead to concepts such as Hartree and Hartree-Fock interactions, and to higher-order scattering processes which involve dissipation.

9. Some Considerations on the Density Matrix

To use the density matrix to describe transport in quantum structures, we must find a set of governing equations that describe the behavior of particles in the appropriate structures, including the role of scattering, and the all important effects of the boundary conditions. In this sense, the situation is exactly the same as for the classical modeling problem, except here our equations have a quantum origin, as discussed above. A key issue in any modeling problem is the computation of the current, and with the density matrix this may well be the most difficult part of the problem. In this section, we will attempt to illustrate how this is done by way of example, and use some particularly simple cases, such as Ohm's law, before proceeding to more complex calculations for barriers. We will do this in somewhat more detail with the density matrix, than with the other approaches to be discussed in subsequent sections, because the basic techniques do not differ in most cases, but are important to understand. In general, dissipative transport is treated through a perturbative treatment, as introduced in the preceeding section through the interaction representation. This approach has had a rich history.[183,275–279] The density matrix approach has been used to compute transport in the presence of a uniform, high electric field in order to sense the changes arising from the quantum distribution function.[280–285] Monte Carlo techniques have been developed to study transport in this regime,[283,286] especially for studying transient transport.[287–289]

It is possible to work with a density matrix that is described in the momentum representation; e.g., the density matrix is a function of two momentum variables rather than the two position variables in (22). Nevertheless, we will describe only the case in the position representation, for which the equation of motion is given by (22). This latter equation is a partial differential equation permitting specification of conditions at boundaries as well as of an initial condition at some time t_0. For those situations where current flows and dissipation is an issue, such as in actual devices, (22) must be modified to include a term representing the decay of the density matrix due to scattering processes. We will spend considerable time with this last term. As we discussed earlier, the problems with devices involve open systems. If the boundaries are characterized by a local equilibrium, as is often the case, the form of the density matrix may be obtained from the semi-classical Fermi-Dirac distribution through the inverse transform of (28). To date, the description of transport in devices has been confined to the cases where the particles are free in two directions (one-dimensional transport through the devices), which for specificity we take as the y and z directions. Hence, we can set $y = y'$ and $z = z'$, so that for parabolic bands,

$$\rho(x,x') = N_c \frac{\lambda_d}{(x-x')\Gamma(\frac{3}{2})} \int_0^\infty d\mu \frac{\sin\left[\mu^{1/2}(x-x')/\lambda_d\right]}{1+\exp(\mu-\mu_F)} , \tag{86}$$

where $\mu = E/k_BT$, $\mu_F = E_F/k_BT$, $\lambda_d = \sqrt{\hbar^2/2mk_BT}$ (which differs from the thermal Debye length introduced earlier), all energies are measured from the conduction band edge, and $N_c = \Gamma(\frac{3}{2})/(2\pi^2\lambda_d^3)$. In the limit that $x \to x'$, the terms following the effective density of states become just the Fermi-Dirac integral $F_{1/2}(\mu_F)$. There are two limiting cases that may be easily analyzed. In the high temperature limit, where Boltzmann statistics apply, (86) reduces to the Gaussian

$$\rho(x,x') = N_c \exp\left[\mu_F - \frac{(x-x')^2}{4\lambda_d^2}\right] . \tag{87}$$

For a material such as GaAs, λ_d at room temperature is 4.7 nm and the effective density of states is about $4.4 \times 10^{17}\ cm^{-3}$. For a nominal density of $10^{17}\ cm^{-3}$, the Fermi level lies about $1.5k_BT$ below the conduction band edge, which is borderline on the applicability of the non-degenerate form.

The second case of interest is the low temperature limit, the so-called quantum limit, at $T = 0$. Here, the Fermi level must lie in the conduction band for any reasonable carrier density, and (86) becomes

$$\rho(x,x') = \left[\frac{k_F^3}{\pi^2}\right] \frac{j_1[k_F(x-x')]}{k_F(x-x')} , \tag{88}$$

where $j_1(r)$ is a spherical Bessel function. In the limit $x \rightarrow x'$, the second fraction becomes simply $\frac{1}{3}$. One of the earliest applications in which (88) appeared was a discussion by Bardeen[290] in which he showed that the electron density profile a distance x from an infinite potential barrier was given by

$$n(x) = n_0 \left[1 - \frac{3}{2} \frac{j_1(2k_F x)}{k_F x} \right] \ , \ x > 0 \ , \tag{89}$$

and zero for $x < 0$. Oscillations in the density occur as a result of the spherical Bessel function in (88). These oscillations depend upon the Fermi wave vector, which in turn is a function of the density. These oscillations do not occur in the non-degenerate limit of (87). A more exact calculation that interpolates between these two limiting cases can be expected to show a gradual damping of the oscillatory behavior as the temperature is raised. This is shown in Fig. 12, where the density matrix is plotted (for a homogeneous system) as a function of the off-diagonal, or correlation, distance. It may be seen that the oscillations are quite weak already at 77 K, and do not appear at 300 K. At higher densities, the oscillations will exist to higher temperatures.

The above solutions provide some indication of what the density matrix coordinate representation profiles are for the cases most closely related to the standard classical equilibrium distribution functions. It may be anticipated, purely on physical grounds, that a problem examined using the classical distribution function in momentum space would yield the same physical results, with respect to the observables, as obtained in the coordinate representation. Consider, for example, the fact that the Boltzmann distribution $\exp\left[-V(x)/k_B T\right]$ (recall that V is an energy—the potential energy). If we introduce a potential step of amplitude $k_B T \ln(10)$, classical theory leads to the conclusion that the density is reduced by a factor of 10 (assuming of course that we are talking about *allowed* states in both regions). The same basic result is obtained for the density matrix, but if the temperature is low, so that the above oscillations are not damped, there is a more complex transition region between the two asymptotic potential values. This is shown in Fig. 13 for GaAs at $T = 0$ and a doping of $10^{17} cm^{-3}$ in the region with the potential and an order of magnitude higher in the region in which the potential is $V(x) = 0$.

A comparable situation with respect to potential energy and density occurs when, instead of solving a non-self-consistent equation with a barrier, we solve the self-consistent equation for the spatially varying density. This is shown in Fig. 14 for the same doping parameters as Fig. 13. The solution region extends for 100 nm on either side of the *abrupt change in the doping*; e.g., the potential is created by the abrupt change in doping profile. Again, the solution is at $T = 0$. No attempt is made to develop the origin of the carriers, rather the regime is chosen because the detailed values of the density can be obtained independently. Here, the solution to the equation of motion of the density matrix is provided self-consistently by also solving the Poisson equation

$$\nabla \cdot [\varepsilon(\mathbf{r})\nabla V(\mathbf{r})] = -e^2 [\rho(\mathbf{r},\mathbf{r}) - N_D(\mathbf{r})] \ . \tag{90}$$

There are a number of interesting features about the above figures. First, if we concentrate on the shape of the density matrix in the uniform field regions (away from the transition), the ripples in the density matrix indicate that the sharper Fermi edge structure plays a pronounced role as the temperature is lowered, as was pointed out above. In addition, the curvature of the density matrix at the main diagonal, but in the direction normal to the main diagonal, is more pronounced in the region with lower potential. This directly relates to the (average) kinetic energy of the density through[291]

$$E(\mathbf{r})n(\mathbf{r}) = \lim_{\mathbf{r}' \to \mathbf{r}} \left[-\frac{\hbar^2}{8m} \left(\frac{\partial}{\partial \mathbf{r}} - \frac{\partial}{\partial \mathbf{r}'} \right)^2 \rho(\mathbf{r},\mathbf{r}') \right] \ . \tag{91}$$

This should be compared with the same form in (78). As the temperature is raised, the curvature will be decreased, as the particles spread and thecorrelation extends further. Moreover, increased density results in increased curvature due to the increase of the average energy by the increase in the Fermi energy.

A. Statistics of a Single Barrier

Consideration of the situation with a finite barrier offers similar insights. For example, if the barrier is sufficiently wide (the characterization of sufficiently wide will be discussed below), we expect that the density within the interior of the barrier, far from the potential transitions, is describable by its classical values, again assuming that this region remains classically allowed. This case is shown in Fig. 15 for the same amplitude barrier as above; that is, the density is taken to be $10^{18} cm^{-3}$ in the region far from the barrier and an order of magnitude smaller (classically) within the barrier. However, when the width of the barrier is reduced, there is an increase in the density in the central region, which is also shown in Fig. 15(a). While the explanation of the variation of the density through the barrier region is describable in terms of the internal wave function reflections at the interfaces, along with the normal continuity of the wave functions and their derivatives, a more practical description of the density variation may be found using the quantum potential U_Q introduced in the previous section. By comparing the density matrix solutions with the approximate Boltzmann-like solutions using the quantum potential, it is found that a quite good description of the variation through the barrier can be obtained.[291] This is true for non-degenerate statistics for sure, but a comparable equivalence has not yet been shown to be valid for strongly-degenerate Fermi statistics. Even so, the results found below give strong correlation to this interpretation.

The quantum potential is one of the more interesting concepts that can be probed through simple solutions of the density matrix equation of motion. In this sense, the density matrix can be solved exactly for simple barrier problems, and the quantum potential can then be calculated from the resulting density through the use of e.g. (2). For example, the quantum potential that results for the narrow barrier in Fig. 15 is shown in Fig. 16, along with the non-self-consistent potential barrier. Within the barrier, where the curvature of the density is positive, the quantum potential is negative and the net result is that the effective potential energy seen by the electrons is less than $V(x)$. This results in a density that is larger than the classically expected value. Immediately outside the barrier region, where the density begins to increase (and has a negative curvature), the quantum potential is positive and the density is below its classical value. In fact, in the center of the barrier, the quantum potential for this example has reached a value that almost cancels the barrier potential.

We can extend this simple evaluation approach to the case of a single heterostructure barrier, such as occurs at the interface between GaAs and GaAlAs which is modulation doped with the impurities residing in the latter material. In Fig. 17(a), the charge distribution, potential energy, and quantum potential are shown for a 200 nm region, at $T = 300K$, in which the interface lies at the center of this region. The doping level is taken to be $10^{17} cm^{-3}$ in the GaAs and $10^{18} cm^{-3}$ for the GaAlAs (which is the region $100 < x < 200 nm$). In addition, the composition of the latter material is assumed to be such that a 0.3 eV barrier is created by the offset of the two conduction bands. It is easily seen that there is a reduction in the carrier density in the GaAlAs near the interface, with a resulting formation of an inversion layer in the GaAs adjacent to the interface. The peak in the density in the inversion layer is actually higher than the background doping of the GaAlAs. It should be noted that the applied potential energy difference across the interface has been chosen yield flat-band conditions, and is equal to the height of the barrier plus the built-in potential of the "junction." For the situation where the charge depletion occurs in the GaAlAs at the heterobarrier interface, the amount of this depletion is such that the curvature of the potential energy within the vicinity of the barrier is approximately constant. As a consequence, when the height of the heterobarrier increases there is an increase in the width of the depletion zone on the GaAlAs side of the structure. Under flat-band conditions, where the net charge distribution is zero, there is a corresponding increase in the charge on the GaAs side. The quantum potential is negative on the GaAlAs side of the junction and tends to give a charge density that is actually larger than the classical value that would be expected. This also has the result of giving a slightly lower value on the opposite side of the junction than what would normally be expected. The small region of negative quantum potential on the GaAs side of the junction is a consequence of oscillations in the density that arise from the Fermi distribution.

B. Multiple-Barrier Structures

The simplest multiple-barrier structure is the double-barrier resonant-tunneling diode, which was introduced previously as one of our proto-typical quantum devices. The characteristic feature of the multiple-barrier structures is the existence of quasi-bound states between each pair of barriers (if there are more than two barriers). The density between the barriers, as well as within the barriers themselves, depends upon the potential height of the barrier, the configuration (spatial variation) of the barrier, the doping levels, as well as the size of the regions between the barriers. The value of the quantum potential in the region between the barriers is approximately equal to the energy of the lowest quasi-bound state, relative to the bottom of the conduction band.[292] It has been shown previously that, in the absence of current normal to the barriers, the total energy in this region is given by $E = V(x)+U_Q(x)$.[257] We show this by considering a region in which two $5nm$ barriers, $0.3eV$ high, separated by $5nm$, are placed in the central part of the simulation region ($200nm$ long). The doping is taken to be $10^{18}cm^{-3}$, except in a central $40nm$ region in which it is reduced by two orders of magnitude. In Fig. 18, we plot the density distribution, the donor concentration, the quantum potential, and the self-consistent potential. It may be seen that the peak density in the center of the quantum well rises to a value that is about 40% of that in the heavily-doped regions. The quantum potential, as expected from the above discussion, is negative in the barriers and is positive in the quantum well. For this results of this figure, the quantum potential rises to approximately $84meV$, which is quite near the computed value of the quasi-bound state. (If the barriers are reduced to $200meV$, the quantum potential peak is reduced to about $70meV$.) The connection of the quantum potential to the quasi-bound state is an important feature of modeling with the density matrix, as it allows an easy evaluation of bound energy levels in complicated structures.

A second calculation may be used to further examine the expected results. Here, the double-barrier structure above is placed into a $40nm$ quantum well, with the depth of the well a variable. As the depth of this well was increased, the quantum potential value in the central well remained independent of position, but increased slightly in value due to the extra confinement effect of the external (large) quantum well. When this larger quantum well was $150meV$ deep, the quantum potential, and hence the quasi-bound level, increased to about $94meV$. For this condition, the various observables, and the quantum potential are shown in Fig. 19. It may be seen from this figure that the density between the two barriers has increased. Such an increase has at least two origins: i) the increased density on either side of the barriers, due to the confining effect of the larger well, and ii) the lowering of the lowering of the quasi-bound state relative to the Fermi level in the heavily-doped regions. The quantum well itself is delineated in the figure by the subsidiary offset of the potential energy at the boundaries of the large quantum well.

Frensley[293] has used the single-particle density matrix to study a double-barrier resonant tunneling diode, and also looked at the case with more built-in barriers. The partial differential equation for the density matrix was solved using finite-difference techniques similar to those used to solve in conventional semiocnductor devices. For the simulations, he treated the boundaries as ohmic contacts. He finds that the methods of solution are easier than the Wigner distribution fucntion (discussed in the next section), but are more complicated to interpret, as may be inferred from the above discussion. He assumed 50 *nm* barriers and well, and used a heterostructure offset of 0.39 *eV*. This work is the first step toward directly calculating the properties of real quantum devices, and points out the importance of such simulations to gain insight into the operation of the devices. Nevertheless, incorporation of the ohmic boundary conditions greatly complicated the numerical simulation algorithm, and the system was subject to the growth of numerically unstable modes. This continues to be a problem. However, we point out below that within the Wigner formulation there exists a methodology to overcome this problem.

Groshev[294] has also used the density matrix to simulate the Coulomb blockade (single-electron tunneling) regime of the resonant tunneling structure. In this case, he used a three-dimensionally configured structure in an attempt to identify lateral modes and fine structure in the tunneling current. However, he reduced the problem to a pseudo-hopping formalism which did not need to carefully study the spatial charge distribution self-consistently, nor did it carefully examine the effects of the boundary conditions.

10. Dissipation and Current Flow

One conclusion we can draw from the above considerations is that, for both the self-consistent and non-self-consistent solutions of the potential, the solutions for the density and the potential sufficiently far from the interface are basically the same as that expected using the classical Boltzmann equation. When current flows through the structure, the semi-classical approach is usually pursued either by drift-diffusion or by hydrodynamic approaches, or through more extensive simulation of the Boltzmann equation through ensemble Monte Carlo procedures.[11,180] Here, we want to begin to understand how current transport in quantum structures can be treated via the density matrix equations of motion. For cases where the ends of the "device" are heavily-doped n^+ regions, boundary conditions on the numerical procedures are formulated to assure that the numbers of particles entering the "cathode" end of the device is equal to the number of particles leaving the "anode" end of the device (for electron flow). An alternative schema is to adopt procedures which incorporate sufficient dissipation in these boundary regions to thermalize the carriers to a near-equilibrium distribution. This latter schema should yield the same results for the charge and potential within the active portion of the device, but the major problem is how to deal with the dissipation processes within the simulations. To date, only very approximate schemes have been adopted, but it must be emphasized that some procedure for treating the dissipation must be adopted if transport in such devices is to be discussed sensibly. A low-order (Born approximation) perturbation treatment of phonon interaction in a homogeneous system has been treated with the density matrix by Argyres.[295] Similar approaches have been used for heterostructures.[296] A more extensive renormalized phonon treatment has also been treated for homogeneous systems.[297] We will follow a somewhat different approach here, in order to try to develop an effective approach for detailed device simulations.

Following Caldeira and Leggett,[109,298] we consider a system A interacting with a second system R (which is taken normally to be the reservoir) and described by the Hamiltonian $H_T = H + H_R + H_I$, where the latter three terms describe system A, the reservoir and the interaction between the two. Clearly, we can follow the approach of Sec. 5.c, and define the projection operator as being a trace over the reservoir variables, so that the reduced density matrix is described by[12]

$$\rho_d(\mathbf{r},\mathbf{r}',t) = P\rho(\mathbf{r},\mathbf{r}',\mathbf{X},\mathbf{X}',t) = Tr_{\mathbf{X},\mathbf{X}'}\{\rho(\mathbf{r},\mathbf{r}',\mathbf{X},\mathbf{X}',t)\}\ , \qquad (92)$$

where $\mathbf{X},\mathbf{X}'$ are the coordinates of the reservoir. In this approach, we do not need detailed information about the reservoirs, only information about their interactions with, and influence on, the electron system in the active device region. The method that is normally invoked is to develop a perturbation theory for the interaction and dissipation that follows from both internal dissipative processes and boundary effects (one such is surface-roughness scattering; another appears as changed boundary conditions which will be heavily used in the next section on Wigner functions). The equation of motion for the density matrix (22) is rewritten to include a scattering contribution, which must be evaluated, which is the next task.

In determining the form of the perturbation theory result, one is faced with a variety of approaches to take. One approach which encourages an intuitive prescription is to simply determine how standard Boltzmann scattering terms would look in the appropriate coordinate representation. While this will lose information regarding short-time events and build-up of the scattering process, much insight is gained into the form of scattering to low-order in perturbation theory. In the Boltzmann picture, when Fermi statistics can be safely ignored, the scattering rate is simply

$$\left.\frac{\partial f(\mathbf{r},\mathbf{k})}{\partial t}\right|_{diss} = -\frac{1}{4\pi^3}\int d^3\mathbf{k}'\left\{f(\mathbf{r},\mathbf{k})W(\mathbf{r},\mathbf{k},\mathbf{k}') - f(\mathbf{r},\mathbf{k}')W(\mathbf{r},\mathbf{k}',\mathbf{k})\right\}\ , \tag{93}$$

where the scattering processes are assumed to occur locally in space and $W(\mathbf{r},\mathbf{k},\mathbf{k}')$ is the standard Fermi-golden-rule transition probability. Our approach then assumes that $f(\mathbf{r},\mathbf{k})$ replaced in the quantum treatment by a Wigner distribution and the inverse of the Weyl transform (28) is used as

$$\rho(\mathbf{R}+\frac{\mathbf{s}}{2},\mathbf{R}-\frac{\mathbf{s}}{2}) = \frac{1}{4\pi^3}\int d^3\mathbf{k} f_W(\mathbf{r},\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{s}}\ , \tag{94}$$

where the reduced coordinates (26) are used. After some simple manipulations, the scattering term (93) becomes

$$\left.\frac{\partial \rho(\mathbf{R}_\pm)}{\partial t}\right|_{diss} = -\frac{1}{4\pi^3}\int d^3\mathbf{k}\int d^3\mathbf{k}'\left\{f_W(\mathbf{r},\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{s}}W(\mathbf{r},\mathbf{k}',\mathbf{k})\left[1-e^{-i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{s}}\right]\right\}\ , \tag{95}$$

where $\rho(\mathbf{R}_\pm) = \rho(\mathbf{R}+\frac{\mathbf{s}}{2},\mathbf{R}-\frac{\mathbf{s}}{2})$.

The structure of the scattering term in the coordinate representation may now be obtained from (95), at least as the scattering is derived to lowest order in the Boltzmann scattering framework. For scattering which is mainly local in space, the difference coordinate $\mathbf{s}$ is small, and the second term in the square brackets can be expanded in a Taylor series, retaining only the leading non-vanishing term, for which

$$\left.\frac{\partial \rho(\mathbf{R}_\pm)}{\partial t}\right|_{diss} = -\frac{i\mathbf{s}}{4\pi^3}\cdot\int d^3\mathbf{k}\int d^3\mathbf{k}'\left\{f_W(\mathbf{r},\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{s}}W(\mathbf{r},\mathbf{k}',\mathbf{k})(\mathbf{k}'-\mathbf{k})\right\}\ . \tag{96}$$

It may be recognized, from standard semi-classical treatment of scattering processes,[263] that the integration over $\mathbf{k}'$ produces a momentum relaxation rate. Introducing (28), and integrating once by parts on the s coordinate, the dissipation rate can be rewritten as

$$\left.\frac{\partial \rho(\mathbf{R}_\pm)}{\partial t}\right|_{diss} = -\mathbf{s}\cdot\frac{\partial}{\partial \mathbf{s}}\int d^3\mathbf{k}\int d^3\mathbf{s}'\rho(\mathbf{R}_\pm)e^{-i\mathbf{k}\cdot(\mathbf{s}-\mathbf{s}')}\Gamma(\mathbf{r},\mathbf{k})\,. \tag{97}$$

We note that the integration over k incorporates the rapidly varying exponential, which in turn is approximately a delta function on the difference in the nonlocal variables, so that one really achieves an average value of the momentum relaxation rate, and

$$\left.\frac{\partial \rho(\mathbf{R}_\pm)}{\partial t}\right|_{diss} = -\Gamma(\mathbf{r})\mathbf{s}\cdot\frac{\partial}{\partial \mathbf{s}}\rho(\mathbf{R}_\pm)\,. \tag{98}$$

This term can now be added to the Liouville equation (22) on the right-hand side (after multiplying by $i\hbar$, of course). This form of the dissipative term has also been discussed by Dekker,[299] as well as in the work cited earlier.[298] Density matrix algorithms for modeling devices have been reported that incorporate the dissipative contributions developed here.[292] Because of numerical difficulties at higher bias levels, modifications to the scattering term were introduced that go beyond the approximations introduced in arriving at (98). We turn to these next.

One may note that it is possible to rewrite the dissipative term by noting that the current density itself is a function of the non-local variable, through the fact that

$$\mathbf{j}(\mathbf{r},\mathbf{r}') = -\frac{i\hbar}{2m}\left(\frac{\partial}{\partial \mathbf{r}} - \frac{\partial}{\partial \mathbf{r}'}\right)\rho(\mathbf{r},\mathbf{r}') = -\frac{i\hbar}{m}\frac{\partial \rho(\mathbf{R}_\pm)}{\partial \mathbf{s}}\,, \tag{99}$$

where we have used the coordinate transformation (26). With this form of the current, the dissipation term can now be rewritten as

$$\left.\frac{\partial \rho(\mathbf{R}_\pm)}{\partial t}\right|_{diss} = -\frac{im}{\hbar}\Gamma(\mathbf{R})\mathbf{s}\cdot\mathbf{j}(\mathbf{R}+\frac{\mathbf{s}}{2},\mathbf{R}-\frac{\mathbf{s}}{2})\,. \tag{100}$$

The current term can now be rewritten in terms of a velocity, under the assumption that this local velocity incorporates any diffusion effects directly, which is a local quantity of the average (center-of-mass) coordinate, or $\mathbf{j}(\mathbf{r},\mathbf{r}') \approx v(\mathbf{R})\rho(\mathbf{r},\mathbf{r}')$. Further, it may be noted that the relaxation rate plays a nearly semi-classical role in near equilibrium, in that the potential drop between any two points can be described in terms of this dissipation and the velocity, much as a quasi-Fermi level is introduced, or

$$E_F(\mathbf{r}) - E_F(\mathbf{r}') = -\int_{\mathbf{r}}^{\mathbf{r}'} d^3\mathbf{u} \cdot \mathbf{v}(\mathbf{u}) m\Gamma(\mathbf{u}) \,. \tag{101}$$

For small values of s, this latter equation is approximately -$\mathbf{s} \cdot \mathbf{v}(\mathbf{R}) m\Gamma(\mathbf{R})$, and the dissipative term finally becomes

$$\left.\frac{\partial \rho(\mathbf{R}+\frac{\mathbf{s}}{2}, \mathbf{R}-\frac{\mathbf{s}}{2})}{\partial t}\right|_{diss} = \frac{i}{\hbar}\left[E_F(\mathbf{r}) - E_F(\mathbf{r}')\right] \rho(\mathbf{R}+\frac{\mathbf{s}}{2}, \mathbf{R}-\frac{\mathbf{s}}{2}) \,, \tag{102}$$

so that the dissipative term has precisely the same form as the potential driving term, but with the potentials replaced by the (quasi-) Fermi levels. This has the result of creating a dissipative term on the right-hand side of (22) that is still real. Side-by-side calculations at low values of bias yield identical results.

The manipulations associated with the above discussion were predicated upon finding a means of computing the current. In fact, a definition of the current was introduced in arriving at the final form (102). In the computations that follow, an assumption is made that the carriers at the upstream boundary are in local equilibrium and that the distributions are either a displaced Maxwellian or a displaced Fermi-Dirac distribution. This implies that, at the upstream boundary, the zero current quantum density matrix is replaced by $\rho(\mathbf{r}, \mathbf{r}') \exp\left[\frac{im}{\hbar} \mathbf{v}(\mathbf{R}_{boundary}) \cdot \mathbf{s}\right]$. Since the current is introduced as a boundary condition to the problem, a prescription is necessary to find its value. This leads to an auxiliary condition.

To find the value of the current used in the Liouville equation for the density matrix, moments of the revised form of (22) are taken in the same sense as those of Sec. 8. Under time-independent conditions, the momentum balance equation (77) can be written as

$$2\frac{\partial E}{\partial \mathbf{R}} + \frac{\partial V}{\partial \mathbf{R}} \rho(\mathbf{R}) + m\Gamma(\mathbf{R}) v(\mathbf{R}) \rho(\mathbf{R}) = 0 \,, \tag{103}$$

where E is the total energy, kinetic plus potential, and not the energy per particle as indicated in (78). Under the assumption of current continuity, and the condition that the term $v(\mathbf{R})\rho(\mathbf{R})$ is independent of distance (Kirchoff's current law), ***and the condition that the total energy of the entering and exiting carriers are equal,*** current is then obtained from the condition

$$j(x) = -\frac{\int_0^L dx' \frac{\partial V}{\partial x'} \rho(x')}{\int_0^L dx' \Gamma(x')} , \tag{104}$$

where we have restricted the considerations to one spatial dimension.

The simplest type of calculation to deal with is that of a free particle. For this case, with current introduced as a boundary condition through the above approach, the resulting density matrix is complex. The real part is symmetric in s, the distance from the principle diagonal. On the other hand, the complex part (which leads to the current) is anti-symmetric about the principal diagonal. In Fig. 20, we plot the real and imaginary parts of the density matrix for a 200nm region of GaAs, doped to a level of $10^{17} cm^{-3}$. A small bias of 10meV has been applied across this length (a field of 500 V/cm). It has been assumed that Γ corresponds to a mobility of $2580 cm^2/Vs$,or a scattering time of 0.1 picosecond. The mean carrier velocity is found to be about $1.3 \times 10^6 cm/s$. Increasing the applied bias results in an increase in the velocity and an increase in the kinetic energy of the carriers, the latter of which manifests itself as increased curvature of the imaginary part of the density matrix in the correlation direction (normal to the principal diagonal).

Let us now turn to a non-uniform sample in which the mobility varies with position in the structure. The material is again taken to be GaAs, with the parameters discussed in the previous paragraph for Fig. 20. Here, however, the scattering rate will be greatly increased over the central 2nm of the structure. On the basis of the above discussion, this decrease in the scattering time will result in a sharp drop in the quasi-Fermi level over this region. The density cannot change as rapidly, and has a characteristic "Debye" length over which it changes. The density variation, the quasi-Fermi level, and the potential energy are shown in Fig. 21 for two cases corresponding to an increase of scattering rate by one, and two, orders of magnitude (actually the background scattering rate also is varied in the second). In Fig. 22, the opposite results, for a central region with lower scattering rate, are shown. For the case where the cladding region has a higher mobility (lower scattering rate), most of the potential drop is across the central, low mobility region (Fig. 21). Conversely, when the cladding region has a lower mobility than the central region, most of the potential drop is across the central region (Fig. 22). Of course, these results are clearly expected from classical considerations. What is not usually appreciated is that both cases lead to strong variations in the local, self-consistent density and in the resulting quasi-Fermi levels. It is these regions, where the density varies considerably from the normal doping levels that we expect to see the largest quantum-induced changes in the results.

We now turn to more device-like simulations. A single-barrier tunneling structure is considered. The basic structure is taken to be GaAs with nominal doping of $10^{18} cm^{-3}$. A central $30 nm$ region is assumed to be unintentionally doped. A barrier with $0.3 eV$ height, and $10 nm$ width, is placed in the center of the lightly-doped region. The scattering time is assumed to be constant throughout the structure at a value of 0.1 ps. In Fig. 23, the potential and charge density are shown for a variety of applied bias levels, ranging from 0 to 0.4 eV. The bias is applied at the collector boundary (the zero of potential is maintained at the source end). A typical potential well forms on the source side of the barrier as the bias is increased (we will see similar behavior in the DBRTD of the next section). In all cases, the potential decreases essentially linearly across the barrier, which indicates that there is little charge accumulation (or depletion) in the barrier itself. On the other hand, there is significant charge accumulation on the source side of the barrier at the higher bias levels. In Fig. 24, the quasi-Fermi level (relative to the equilibrium Fermi level) is shown for the same bias levels. Most of the change in the quasi-Fermi level occurs *within* the barrier, and on the collector side of the barrier. It matches the applied bias well at the boundaries, as required for consistency. The current-voltage curve for this structure shows the expected exponential behavior, although the charge accumulation modifies slightly the pure exponential behavior of simple theory. The shape of this latter curve is also sensitive to the exact dimensions of the tunneling barrier, since this modifies the fraction of the potential which is dropped across the barrier.

11. Further Considerations on the Density Matrix

A. An Alternative Approach to the Density Matrix

An alternative approach to dealing with transport through use of the density matrix has been proposed by Iafrate and Krieger.[279] These authors considered both uniform and nonuniform fields in an approach that allowed them to examine transport on a short-time scale. For example, in a discussion of electrons interacting with impurities and a homogeneous time-dependent electric field in the momentum representation, they used first-order time-dependent perturbation theory to obtain

$$\rho(\mathbf{k},\mathbf{k}',t) = \frac{1}{i\hbar}\int_{t_0}^{t} dt' \langle \mathbf{k}' | H'(t') | \mathbf{k} \rangle \left[\rho(\mathbf{k},\mathbf{k},t) - \rho(\mathbf{k}',\mathbf{k}',t)\right] \times \exp\left\{ \frac{i}{\hbar} \int_{t}^{t'} dt'' \left[E(\mathbf{k}') - E(\mathbf{k})\right] \right\} , \tag{105}$$

which contains several important features. First, the basis functions used in obtaining this equation are the instantaneous eigenfunctions of the Hamiltonian; i.e.,

$$H\psi = \frac{1}{2m}(\mathbf{p} - e\mathbf{A})\cdot(\mathbf{p} - e\mathbf{A})\,\psi = E_0\psi , \tag{106}$$

where

$$E_0(\mathbf{K},t) = \frac{\hbar^2}{2m}\left(\mathbf{K} - \frac{e\mathbf{A}}{\hbar}\right)\cdot\left(\mathbf{K} - \frac{e\mathbf{A}}{\hbar}\right) , \quad \mathbf{k}(\mathbf{K},t) = \mathbf{K} - \frac{e\mathbf{A}}{\hbar} , \tag{107}$$

and

$$E(\mathbf{K},t) = E_0(\mathbf{K},t) + \langle \mathbf{K} | H' | \mathbf{K} \rangle . \tag{108}$$

Second, the off-diagonal elements of the energy in the last equation and the density matrix itself are expressed only in terms of the perturbed diagonal elements. Thus, initial conditions are not subject to the usual condition of an initial equilibrium state. The first-order perturbed density matrix in the momentum representation was used to obtain the time rate of change of the diagonal components themselves. The exponential term, which is field- and time-dependent, permits a discussion of short-time behavior as well as the long-time energy conserving delta-function behavior representative of the Fermi golden rule. The intra-collisional field effect enters at this point. These authors also obtained the Liouville equation in an accelerated Bloch representation, where it was demonstrated that the major result is to introduce a term in the perturbing matrix elements that connects states of the same general momentum K value, but with different band indices, thus leading to a contribution from interband tunneling. It is expected that, as the Wigner function formalism provides an initial form for the modification of the equation of motion for the density matrix, this approach will find application to the examination of transport on the short-time scale.

The situation with nonuniform fields is different, and was treated by these authors accordingly. Here, they examine solutions specifically to the Liouville equation in the Wannier representation; a representation in localized wave functions in real space. This representation bears close resemblance in form to the density matrix in the coordinate representation. The transformation to the Wannier representation is

$$\langle \mathbf{x}|\mathbf{k}\rangle = \sum_n \langle \mathbf{x}|n\rangle \langle n|\mathbf{k}\rangle = \sum_n \psi_n(\mathbf{x})\phi_n(\mathbf{k}) , \tag{109}$$

where $\phi_n(\mathbf{k})$ is the Fourier transform of the coordinate wave function $\psi_n(\mathbf{x})$. Here, $\langle \mathbf{x}|\mathbf{k}\rangle$ is the Wannier function satisfying the orthonormality conditions

$$\sum_{\mathbf{k}} \langle n|\mathbf{k}\rangle \langle \mathbf{k}|m\rangle = \delta_{nm} , \sum_{\mathbf{x}} \langle n|\mathbf{x}\rangle \langle \mathbf{x}|m\rangle = \delta_{nm} , \sum_m \langle \mathbf{k}|m\rangle \langle m|\mathbf{k}'\rangle = \delta(\mathbf{k}-\mathbf{k}') . \tag{110}$$

The density matrix may then be transformed into the Wannier representation as

$$\langle \mathbf{x}\,|\rho|\,\mathbf{x}'\rangle = \sum_{n,m} \langle \mathbf{x}|n\rangle \langle n|\rho|m\rangle \langle m|\mathbf{x}'\rangle , \tag{111}$$

where $\langle n|\rho|m\rangle$ is the density matrix in the Wannier representation (it should be pointed out this this would be true for any arbitrary set of coordinate representation wave functions, but here we have asserted that these wave functions $\langle \mathbf{x}|m\rangle$ are the localized Wannier wave functions). Since the basis functions satisfy the Schrödinger equation, but with the proviso that the energy is now a matrix quantity, the Wannier functions satisfy

$$-\frac{\hbar^2}{2m}\nabla^2 \langle \mathbf{x}|n\rangle = \sum_m E(m,n) \langle \mathbf{x}|m\rangle . \tag{112}$$

Then, the single-band equation of motion (with the short-hand $\rho_{nm} = \langle n|\rho|m\rangle$) is

$$\begin{aligned} i\hbar\frac{\partial \rho_{nm}}{\partial t} &= \sum_{m'} [E(n,m')\rho_{m'm} - \rho_{nm'}E(m',m) \\ &\quad + V_{mm'}\rho_{m'm} - \rho_{nm'}V_{m'n}] . \end{aligned} \tag{113}$$

This latter equation was discussed by Krieger and Iafrate[279] with a band index included.

As the density matrix in the coordinate representation with a finite lattice is indexed by the lattice points, it may be regarded as a pseudo-coordinate representation. It is intriguing to compare the structure of (113) to the Liouville equation itself in the coordinate representation. First, we note that the energy operator in the coordinate representation is a second derivative and, in establishing its contribution, difference equations are implemented which generally involve at most nearest and second-nearest neighbors. The potential energy contributions in this representation do not involve a summation over the coordinates, but are local in the site. Thus, (113) differs from the Liouville equation in the treatment of the potential terms, just as arises in the Wigner representation. The density matrix in the Wannier representation, particularly for well localized values of the momentum, will likely involve significant contributions from lattice sites well removed from that of interest. On the other hand, where the density is well localized, such as in a quantum well, the summation may well involve only a few nearest-neighbor terms. As usual, the choice of the best representation is one of convenience to the problem at hand.

Rather than treat the perturbation of the density matrix in a specific representation, another approach[300] is to deal with perturbation about the equilibrium density operator itself. Here, we represent $H_0(t)$ as the unperturbed Hamiltonian, so that the field and corresponding perturbations are initialized at $t = 0$. We represent the perturbing operator as $V(t)$, in analogy with (35). Then, to second order in the expansion of the unitary perturbation operator, the density matrix can be written as

$$\rho(t) = \rho_0(t) + \rho_1(t) + \rho_2(t) \,, \tag{114}$$

where

$$\rho_0(t) = U_0(t, t_0)\rho_0(t_0)U_0(t_0, t) \,,$$

$$\rho_1(t) = -\frac{i}{\hbar}\int_{t_0}^{t} dt' \, [V(t', t), \rho_0(t)] \,,$$

$$\rho_2(t) = -\frac{1}{\hbar^2}\left\{\int_{t_0}^{t} dt' \int_{t_0}^{t'} dt'' V(t'', t)V(t', t)\rho_0(t) + \rho_0(t)\int_{t_0}^{t} dt' \int_{t_0}^{t'} dt'' V(t', t)V(t'', t)\right.$$

$$-\int_{t_0}^{t} dt' V(t',t)\rho_0(t)\int_{t_0}^{t} dt' V(t',t)\Bigg\} \ . \tag{115}$$

Here, we introduced the two-time potential

$$V(t',t) = U_0(t',t)V(t)U_0(t,t') \ , \tag{116}$$

and the unitary operator

$$U_0(t',t) = \exp\left[-\frac{i}{\hbar}\int_{t}^{t'} H_0(t'')dt''\right] \ . \tag{117}$$

This approach is quite useful and has led to the development of quasi-equilibrium density operators in which the equilibrium density matrix is parameterized much like a drifted Maxwellian. The approach, however, is largely limited to those problems in which linear response around a quasi-equilibrium steady state is possible. It is not at all clear that this non-renormalized approach is applicable in far-from-equilibrium situations.

B. Differential Capacitance

The measurement of the differential capacitance as a function of the bias potential permits a reconstruction of the density profile in the heterostructure from a standard textbook formula

$$n(x) = \left[\frac{d}{dV}\left(\frac{1}{C^2}\right)\right]^{-1} \ , \tag{118}$$

where here the position x is to be interpreted as an average position defined through the capacitance $x = \varepsilon/C$, and C is the specific capacitance per unit area. The extrapolated value of C^{-2} (extrapolated back to the intercept of the voltage axis) gives a value of potential that may be interpreted as the *offset* voltage for a simple heterostructure configuration. This leads us to conclude that a single measurement of capacitance as a function of bias voltage determines both the density and the heterostructure offset voltage.

As a result of the above considerations, the measurements of CV in simple systems have also included a numerical component involving the solution of Poisson's equation for a density distribution computed for a specific heterostructure configuration, from which computed CV characteristics may be obtained. The theoretical structure providing the closest fit between the calculated and measured CV relationships is often pronounced as the one representing the actual heterostructure. The degree to which such a measurement, and the fit to the data, is reliable is dependent upon: (a) the fundamental equations chosen to represent the structure (quantum structures require equations describing quantum transport and not the classical equivalent); (b) the statistics, either Boltzmann or Fermi-Dirac statistics, can influence the actual value of the intercept on the voltage axis to first order; (c) traps and unusual doping contributions (such as planar doping); (d) specific quantum boundary conditions. In short, the most representative calculation is that with the most physics. Those equations least likely to inspire confidence are the usual classical ones that are often used: the drift and diffusion equations. The equations most likely to inspire confidence are those yielding the quantum distribution function, such as the density matrix or the Wigner distribution.

Let us consider a typical calculation to illustrate the point. The quantum approach is to couple the Liouville equation for the density matrix to the Poisson equation and solve the overall structure self-consistently. In the absence of any bias, the CV relationship was obtained for a 200 nm long simulation region, nominally doped to 10^{18} cm^{-3}, and with a 15 nm thick, 300 meV barrier located in the center of a 30 nm non-intentionally doped region. The density and potential variation are shown in Fig. 25. The capacitance is obtained via an approach proposed by Krömer *et al.*[301] From Poisson's equation, the change in the charge density $\delta\rho(x)$ that arises as a consequence of the change in potential $\delta V(L)$ at the end of the sample $x = L$ yields a net total charge change in the structure (the notation corresponds to Fig. 25, where $x = 0$ is taken in the center of the structure)

$$\delta Q(L) = -e \int_{-L}^{L} dx \delta\rho(x) \, . \tag{119}$$

The differential capacitance is then given by

$$C(V) = \frac{\varepsilon}{e} \frac{\delta Q(L)}{\delta V(L)} \equiv \frac{\varepsilon}{e\bar{x}} \, , \tag{120}$$

and this allows the reconstruction of the density from[302]

$$\bar{n} = 2n_{ref}\frac{L_D^2}{\beta}\left(\frac{d\bar{x}^2}{dV}\right)^{-1}, \tag{121}$$

where $\beta = 1/k_BT$, as previously, L_D is the Debye length ($L_D^2 = \varepsilon/\beta e^2 n_{ref}$), and n_{ref} is a reference density usually taken at a boundary where it should be well known.

For a uniform structure, we expect that $\bar{n} = n_{ref}$. In the inset of Fig. 26, we plot the quantity $\bar{x}^2/L_D^2\beta \sim 1/C^2$ as a function of the bias voltage $V(L)$.[303] The density inferred from this procedure (dotted curve) is plotted in the main part of Fig. 26 for the central region of the structure shown in Fig. 25. The straight lines in the inset of the figure illustrate the expectation for a uniformly doped structure in the absence of a heterostructure barrier. The extrapolated bias intercept of the lower straight line is twice the mean energy of the entering carriers. The variation in slope at the center of the figure is a signature of the presence of the barrier. Also shown (solid curve) is the actual charge density distribution obtained from a solution of the density matrix. Clearly, the inferred value of $\bar{n}(\bar{x})$ in the vicinity of the barrier is at best an approximation to the actual charge density, and the minimum inferred density is somewhat greater than the actual minimum density. The asymmetry in the inferred density has been discussed previously.[302] Still, it is significant that the structure and width of the barrier heterostructure can be obtained from this measurement.

IV. Modeling with the Wigner Distribution

The Wigner formalism offers many advantages for quantum modeling. First, it is a phase-space description, similar to the semi-classical Boltzmann equation, which may be seen by comparing (1) and (29) in the limit of small applied fields. Moreover, in the Wigner formalism, scattering is a local (in space) phenomenon.[46] Because of the phase-space nature of the distribution, it is conceptually possible to use the correspondence principle to determine where quantum corrections enter a problem. At the boundaries (the contacts), the phase-space distribution permits separation of the incoming and outgoing components of the distribution, as we will see in the next section, and this permits modeling an ideal contact, and hence an open system.

However, there are problems at the outgoing contact, which will be discussed below. Still, another advantage to the Wigner formalism is that the Wigner function itself is purely real, which simplifies some calculations and the interpretation of the results. By coupling the Wigner equation of motion (30) to Poisson's equation, we can obtain a fully self-consistent approach to modeling various devices.

Wigner functions have been used extensively in the field of quantum optics, where it has been used to describe coherence of optical fields and to describe polarization and transient superposition effects.[304] Quite naturally, this description has been applied to finding quantum-mechanical solutions to the laser master equation (the Fokker-Planck equation[305]) and for describing quantum noise in lasers.[306] More recently, Wigner functions have been applied to optical systems and signals, where they provide a link between Fourier optics and geometrical optics.[307,308] Two-dimensional Wigner optical distributions have been generated,[309] as well as slices of four-dimensional Wigner optical distributions.[310] The Wigner-function description of optical signals has been investigated for use in elementary pattern recognition as well.[311] From this, it may be seen that the Wigner formalism is especially adapted to displaying the roles of phase interference and quantum resonances.

Numerous early work expounded on the virtues of the Wigner formalism for quantum transport with an eye toward the advantages offered by a phase-space representation.[30,35,46,188] Indeed, there were early attempts to model the transport in DBRTDs with the Wigner formalism.[312] In this section, we will review the development of the Wigner function and its equation of motion. The roles of nonlocality and correlation in the function will be discussed. One important point is the need to achieve a correct initial distribution, as the time development of (30) requires a correct initial state.

In Sec. 5b, the Wigner distribution function was described in terms of a density matrix formulated in the difference coordinates (26), and then Fourier transformed on the difference variable s as in (28). It is easily seen from (28) that there is no requirement in the definition of the Wigner function that requires it to be a positive quantity. For this reason, the Wigner function interpretation as a probability distribution must be handled carefully.[313] Nevertheless, it is a quite useful tool in the study of quantum transport.[314] The issue of the non-locality of the Wigner distribution must be examined carefully, as it has value in the regions in which the density matrix, and the wave functions may not have support. Consider Fig. 27, from which it is evident that a distribution defined in the transformed coordinates (26) may be nonzero at positions **R** where the wave functions themselves are zero. This is an important point, and it illustrates the need to very carefully model the solutions to (30). This problem will arise as a need to extend the range of the Wigner function beyond that of the simple device under study.

12. Methods of Solving the Equations

There are many approaches to modeling semiconductor devices, even after the equations of motion have been agreed upon. This is also true with Wigner functions, in that choices of which equations to use for modeling: the steady-state equations or the time-dependent equation of motion (30). In this review, we will concentrate on solutions that are obtained from the time-dependent equation of motion, as many of the physical properties of interest are related to the time evolution of the device behavior. Even then, there are many methods of choosing discretization schemes and numerical solution techniques. The discussion here will focus upon a successful approach used by one of the present authors to study the DBRTD, although other approaches will be discussed as we come upon their influence.

A. The Initial State

A serious consideration of the Wigner formalism is the entry of nonlocal quantum mechanical effects through the inherently nonlocal potential driving function of (31) in (30). In the limit of slow variations, only the linear term coupling the first derivative of the potential, the force, to the derivative of the Wigner function with respect to the momentum remains important, and this is the semi-classical equivalent that arises in (1). If the potential has no terms of higher order than quadratic in the position, this semi-classical term is the only contribution, and, in fact, the Wigner equation of motion (30) will not reproduce the known quantization in a harmonic oscillator potential. A correct quantum-mechanical steady-state solution to the problem may be found by specifying the correct boundary conditions and solving (30) in the absence of the time-derivative term, although there are subtleties in this approach, not the least of which is that it is known to fail in the harmonic oscillator potential.[315] One further problem is that the correct boundary conditions presuppose knowledge of the state of the system both internally and at the boundaries, as the latter are a function of the former through the nonlocality of the internal potential. Thus, knowledge of the boundary conditions presupposes a full knowledge of the solution being sought, even without the equation of motion.

In order to include all orders of quantum corrections, one of two things can be done. The first is to extend the computational domain sufficiently far from the source of the quantum effects that the system is classical, so that a classical distribution can be used at the boundaries. It has been shown that quantum corrections "heal" over several thermal de Broglie wavelengths.[45] In a reasonable GaAs device at 300 K, this length is nearly 100 nm. The second approach is to develop the adjoint equation for the Wigner function that corresponds to the Fourier transform, in the difference coordinate, of (25) in the time domain (as opposed to the inverse temperature domain).[197] We pursue a variant of this approach, by actually computing the density matrix for the initial condition, and then transforming this latter quantity into the Wigner domain.

If the potential approaches a constant, uniform value as $\mathbf{R} \to \pm\infty$ (where the device is supposed to be located near the origin), then the basis states at large distances are plane waves (in the effective mass approximation). In one dimension, we assume that $V(x) = V^-$ for $x < x^-$ and $V(x) = V^+$ for $x > x^+$, where $x^\pm$ describe the transition regions beyond which the potential is uniform. For equilibrium situations, in which $V^+ = V^-$, this gives scattering states incident from the left, with $k > 0$,

$$\psi_k(x) = \frac{1}{\sqrt{2\pi}}[e^{ikx} + r(k)e^{-ikx}] \tag{122}$$

for $x < x^-$, and

$$\psi_k(x) = \frac{1}{\sqrt{2\pi}}t(k)e^{ikx} \tag{123}$$

for $x > x^+$, where $t(k)$ is the transmission coefficient and $r(k)$ is the reflection coefficient. In a similar manner, states incident from the right are defined by

$$\psi_k(x) = \frac{1}{\sqrt{2\pi}}[e^{-ikx} + r(k)e^{ikx}] \tag{124}$$

for $x > x^+$, and

$$\psi_k(x) = \frac{1}{\sqrt{2\pi}}t(k)e^{-ikx} \tag{125}$$

for $x < x^-$. The density matrix is then defined, in the semi-classical equilibrium regions of the boundaries, as

$$\rho(x,x') = \frac{1}{Z}\left[\sum_n \psi_n(x)\psi_n^*(x')f(E_n) + \int dk\psi_k(x)\psi_k^*(x')f(E_k)\right] , \qquad (126)$$

where Z is the partition function, the sum over n is over bound states which may be localized at some point, E_k is the energy of a scattering state, and $f(E)$ is the Fermi-Dirac distribution function.

From an unnormalized basis function, and using translation matrices, an unnormalized state may be computed on the entire domain. These states are normalized by applying scattering theory, in which wave functions in the presence of a scatterer are compared to those in a reference space. These are related through the Lippmann-Schwinger equation. A useful consequence of this equation is that the scattering states satisfy precisely the same orthonormality relations as unperturbed states.[45,316] Each state contributes to the density matrix according to the thermal distribution function $f(E)$. The partition function is found by considering the limit of $x \to x^-$, also a consequence of the normalization conditions. It is defined to be

$$Z^{-1} = 2\beta\sqrt{\pi}e^{\beta V}\lim_{x\to-\infty}\rho(x,x) . \qquad (127)$$

An algorithm for computing the density matrix is thus available.[45] A set of points $\{x\}$ is chosen at which the density matrix is desired to be known. The potential is then taken to be piece-wise constant between the points (approximating a smooth function in the normal finite difference scheme). Energies are randomly sampled according to the Fermi-Dirac distribution function, and each energy gives a wave vector for which a left-incident and a right-incident scattering state is formed. The states are translated and normalized and their contributions added to the density matrix. The density matrix is then normalized with the partition function. This process can be continued by computing a new potential distribution from the new charge distribution using Poisson's equation, and the process iterated for a self-consistent equilibrium state. The resultant density matrix is then Fourier transformed to yield the Wigner distribution function. In fact, this is the procedure that was followed to obtain the equilibrium Wigner distribution for the DBRTD shown in Fig. 7 above. The Wigner distribution is characterized by a thermal distribution far from the barriers. The oscillations near the barriers are a result of the quantum repulsion from the barriers, which causes a depletion of carriers in this region, although this is not a depletion in the normal classical sense. This quantum repulsion is, in a sense, complementary to barrier penetration: just as a nonzero density penetrates a finite distance into a classically forbidden region, a density deficit exists a finite distance into a classically allowed region, so that the total charge in the device is that necessary to maintain overall charge neutrality.

B. Numerical Discretization and Solutions

The Wigner function equation of motion (30) may be evaluated on a two-dimensional discretized grid for the position and the momentum for one dimension. In general, the Wigner function is a continuous function of $2n$ dimensions in n-dimensional space. Here, we will treat primarily the one-dimensional solutions, although other simulations will be discussed later. We consider that the modeled region exists from $0 \leq x\ (= R) \leq L$. The modeled region is divided into a spatial mesh (using a finite-difference approach) of mesh size Δx chosen so that the features of interest, such as potential barriers, are adequately represented by many grid points. Since the potential in a DBRTD (or even a single-barrier diode) varies over a distance of a few monolayers, an appropriate spatial mesh Δx is of the order of a unit cell of the material being utilized, but typically of the order of 0.25 nm. The mesh size for the momentum variable is found by considering the Fourier transform in the Wigner function equation of motion. the discrete Wigner function is periodic in momentum, due to the discretization of the spatial variable, with a period of $\hbar\pi/(\Delta x)$. (For convenience, the wave vector $k = p/\hbar$ is used.) The variation in momentum is thus confined to the lowest zone of this period in momentum space, which is then discretized into a convenient number of meshes, while assuring that the CFW criterion for linear response[317]

$$\frac{\Delta x}{\Delta t} \leq v\,, \tag{128}$$

where v is the highest velocity in the problem, is met for the time evolution.

The Wigner equation of motion (30) is then discretized using a Lax-Wendroff explicit time differencing.[318] This method retains the second-order terms in the Taylor expansion of $f(t$ [illegible] which introduces a second-order spatial difference term into the equation of motion. This second-order term represents an artificial diffusion, which counteracts spurious numerical diffusion that always arises from the first-order terms.[312] This approach has proven to yield very stable solutions for device simulation of the DBRTD,[319,320] and provides a relatively direct approach to iterating the temporal part of the equation. However, other approaches have also been suggested. Ringhofer has suggested using a *spectral* method that also is found to give good convergence.[321] Arnold, and co-workers,[322] have utilized an operator splitting method, in which the time step is split into two portions. If w_n and w_{n+1} are the approximations to the Wigner distribution function $f(t)$ at t_n and $t_{n+1} = t_n + \Delta t$, and we have two operators A and B, then one first solves the iteration (in the range $t_n \leq t \leq t_{n+1}$)

$$\frac{\partial u}{\partial t} = Au \, , \; u(t_n) = w_n \, , \; w_{n+\frac{1}{2}} = u(t_{n+1}) \, , \tag{129}$$

and techniques such as the Lax-Wendroff approach can be used here. Then, the iteration

$$\frac{\partial u}{\partial t} = Bu \, , \; u(t_n) = w_{n+\frac{1}{2}} \, , \; w_{n+1} = u(t_{n+1}) \tag{130}$$

is solved. Obviously, the overall operator present in (30) (the second and third terms) are split to form the operators A and B, representing the spatial gradient term and the potential term, respectively. Arnold and Nier[332] have also suggested a novel particle flow procedure, very similar to a Monte Carlo simulation for transport, to solve directly (30). This approach bears a similarity to an earlier suggestion of a trajectory approach to solving this equation in DBRTDs.[324] However, in our simulation of the DBRTD,[320] adequate convergence was found with the direct Lax-Wendroff approach.

The spatial variations are handled by noting that each point in the discretized space has a characteristic direction, which describes the direction of probability flow away from that point. In a phase-space representation, the velocity which defines propagation is directly proportional to the momentum. For positive momentum, propagation is in the positive x direction. Conversely, for negative momentum, propagation is in the negative x direction. Thus, it may be observed that the discretized Wigner function consists of *slices* of the mesh, over which the momentum is constant. These slices can be viewed as local systems of equations which are coupled by the potential term of (30). It is this view that leads to the operator-splitting method discussed above. However, here we note that for positive momentum, information flows into the slice from the boundary at $x = 0$, and moves toward the boundary at $x = L$. For negative momentum, the opposite direction of flow is present. Recognition of these characteristic directions solves an intrinsic problem with the Lax-Wendroff scheme. A second-order finite-difference term at point x_i involves the points x_{i-1} and x_{i+1}. In the interior of the device, this creates no problem, nor does it create a problem at the incoming boundary where the boundary distribution is specified. The problem occurs at the outgoing boundary, where the exit distribution is not know and cannot be specified without over-constraining the solution to the differential equation. The solution to this dilemma is to use first-order upwind differencing[318] to propagate the function to the outgoing boundary along the characteristic direction, which is

$$\frac{\partial f}{\partial x} = \frac{f(x_i) - f(x_{i-1})}{\Delta x} \,, \; if \; k(x_i) > 0 \,, \tag{131}$$

and

$$\frac{\partial f}{\partial x} = \frac{f(x_{i+1}) - f(x_i)}{\Delta x} \,, \; if \; k(x_i) < 0 \,. \tag{132}$$

The stability of this overall numerical approach has been checked first by using stationary distributions, and following them for picoseconds of time to assure that no numerical instability creeps into the solution. Secondly, a Gaussian wave packet was propagated through a single tunneling barrier until well formed transmitted and reflected wave packets existed; then, the time variable was reversed and the solution propagated until the latter two wave packets recombined into a *single* wave packet. Any spurious contribution from numerical instability would be expected to destroy the phase memory in the problem and to lead to a two wave packet tunneling problem which would have produced four output packets.[325] All of this was found to be sensitive to the actual boundary conditions used in the problem, which are described next. A somewhat different approach, in which the Wigner distribution is expanded in a set of Wannier functions, leading to a lattice Wigner function, has been suggested[326] and leads to similar stability of the simulation.[327]

In Fig. 28, we show a Gaussian wave packet interacting with a DBRTD potential. The barriers in this case are 3 nm thick, and 0.3 eV high, and are separated by a 5 nm quantum well. In Fig. 28(a), the initial wave packet is shown just as it begins to interact with the barriers (the shaded portions of the spatial axis). In Fig 28(b), the majority of the wave packet has reflected from the barrier, while a small portion has tunneled through the barriers. Finally, in Fig. 28(c), most of the wave packet has reflected from the barrier and has propagated away from the interaction region. Of particular note in this latter figure is the rapidly varying oscillating structure along the $k = 0$ axis. This structure contains the important correlation information between the two outgoing (transmitted and reflected) packets. As long as this information is properly retained (and is convergent in the simulation scheme), the system has time-reversal symmetry as required in the absence of dissipation. When this correlation information is damped, or erased, the original wave packet can no longer be recovered by reversing the time propagation.[328] Any dissipative processes will act to reduce this correlation, and thus destroy reversibility, bringing a clearer understanding of what random-phase approximations mean in this system.

C. Boundary Conditions for the Simulation

Simulation of a real device includes some model of the interface between the interior simulation region and the supposed boundary/contact layer. At the very minimum, the external circuit consists of a battery which fixes the potential across the device, and wires which carry the current from the battery to the device. These circuit parameters are usually included in device simulations as boundary conditions. In many models, the contact serves as an infinite reservoir of thermally distributed carriers.[75] This reservoir maintains a fixed distribution at the contact where the particles enter the simulation domain, and impacts the simulation of such open quantum structures.[329] Conditions at the contact/boundary must be consistent with physical reality. If a current is flowing through the device, current continuity requires that an identical current be flowing through the external circuit. This implies that the external circuit, and thus the contact regime, must be characterized by some distribution which relects current flow, such as a shifted Fermi-Dirac distribution.

In the flow field discussion above, there are left and right boundaries, at $x = 0$ and $x = L$, respectively. Mathematical constraints permit us to specify only one boundary, however (the linear term is only first-order in the spatial derivative). As discussed earlier, the model may be thought of as a coupled set of systems, each of which is a slice in momentum space. Each of these momentum slices has a boundary, or contact, from which electrons enter the slice, and they leave the slice from the opposite boundary. The "entering" contact is constrained by the current. Just inside the device a current exists. Because an identical current must be present in the contact, and the distribution within the contact is assumed to be a near-equilibrium shifted thermal distribution, the amount of the shift and therefore the distribution within the contact are known. The distribution within the contact is then matched to the corresponding momentum slice of the interior model region. This procedure must be carried out self-consistently, since the current within the device is a function of the boundary conditions, which in turn are functions of the current within the device. Thus, the boundary distribution is adjusted self-consistently to provide the necessary transfer of carriers into the device on the source ends of each slice.

A second important property of the contacts is that they must remove all carriers which are leaving the device. In the so-called ideal contact, the carriers are perfectly extracted from the device at the outgoing boundary. The non-local potential in (31) creates a problem as there is usually a weak discontinuity in the derivatives of the potential at the boundary. This leads to an artificial reflection of carriers from the boundary region.[312,330] Such artificial reflections mar the concept of an ideal contact. The problem is well-known in general wave propagation, and the solution lies in coupling the adjacent incoming distribution to the outgoing distribution in a manner that cancels these artificial reflections, particularly from the fastest particles which cause the most trouble.[330] The solution for this problem is to couple part of the outgoing wave back into the incoming wave during the time step process (during the time evolution upgrade iteration). If we rewrite (30) as

$$\frac{\partial}{\partial x} f_w(x,p,t) = a(p)\frac{\partial f_w(x,p,t)}{\partial t} + \frac{1}{h}\int dP M(x,P) f_w(x,p+P,t)\,, \quad (133)$$

where $a(p)$ is the inverse velocity, and the potential function has been suitably modified by dividing by $-v(p)$, the boundary condition can be found by a suitable operator normalization procedure.[330,331] This leads to the boundary condition at $x = 0$, where $p > 0$, for incoming waves to be expressed through the temporal adjustment

$$\frac{\partial f_W(0,p,t)}{\partial t} + \frac{1}{h}\int dP \frac{M(0,P)}{a(p) - a(P)} f_W(0,p+P,t) = 0 \ . \tag{134}$$

Similarly, the boundary condition for the incoming particles at $x = L$, where $p < 0$, is given by the temporal adjustment

$$\frac{\partial f_W(L,p,t)}{\partial t} + \frac{1}{h}\int dP \frac{M(L,P)}{a(p) - a(P)} f_W(L,p+P,t) = 0 \ . \tag{135}$$

These two equations constitute an absorbing boundary condition for the Wigner equation of motion that removes artificial reflections from the outgoing boundary. The reflected waves are absorbed at least to second order. These boundary conditions were used in the simulation of Gaussian wave packets illustrated earlier in Fig. 7. In subsequent work, Arnold has recast these boundary conditions into an analytical framework, from which he demonstrated the well-posedness of these equations, as well as showing that they could be considered as members of a heirarchy of possible boundary conditions.[332]

If we are to accurately model heterostructures, then there are also interior boundaries to consider, such as may arise at the interface between two materials, where there are differences in the effective mass. If the bands are parabolic, this does not introduce much of a problem, as it can be handled in a simplified manner by renormalizing the barrier potential (since, in the scattering state basis, the discontinuity in the momentum wave vector is defined by the product of the mass and energy). However, this may not insure current continuity across the interface, which could introduce a source of error. The proper interface characteristics will take this into account.[333]

13. The Double-Barrier Resonant-Tunneling Diode

In this section, we will describe the simulation of a double-barrier resonant-tunneling diode using the approaches outlined above. Our primary emphasis will be on a GaAs/AlGaAs structure with two 5 nm thick, 0.3 eV barriers separated by a 5 nm well. The regions outside the barriers is assumed to be doped to a level of 10^{18} cm^{-3}, although the importance of using a lightly-doped *spacer*-layer adjacent to the barriers will be discussed as well.

A relaxation-time approximation, with a τ appropriate for a mobility of 3000 cm^2/Vs, will be used to simulate the dissipative processes in the device, although we will discuss the impact of detailed modeling of the inelastic processes in the next section. The equilibrium Wigner distribution function for this structure (in the absence of the spacer layers) was illustrated in Fig. 7.

The steady-state $I - V$ curve of the device is calculated by applying an incremental (negative) bias potential to the cathode contact, then solving (30) to steady-state by time iteration procedures, while solving for the local potential through a self-consistent iteration of Poisson's equation.[320] Other approaches have also directly solved for the distribution function and the potential through a purely steady-state solution of the equations.[319] Here, we concentrate on the time-evolution approach. Study of the transients show that large-signal transients decay exponentially, and that steady-state conditions are achieved in a few hundred femtoseconds.[328,334] The time is stepped in small units up to 1.5 ps to insure that steady-state is reached, while checking simultaneously for convergence (at the end of the time evolution) of the distribution function and potential. Once the maximum applied voltage is reached, the potential is then ramped downward again. One absolute check for the overall stability of the solution technique is the recovery of the equilibrium distribution function at the end of the voltage sweep. Numerical instabilities would be integrated to their maximum amplitude during this process, so that the recovery of the equilibrium distribution means that the total error introduced in the process, integrated over many hundreds of ps, is negligible on the scale of the equilibrium Wigner distribution function. From the steady-state conditions, at each applied bias level, the current in the device is calculated.

The resultant I-V characteristics are shown in Fig. 29. The self-consistent internal potential at a bias level of 0.22 V, quite near the peak in the current through the DBRTD, is shown in Fig. 30. It is clear that only about 1/3 of the total potential is dropped across the barrier region of the device. A majority of the remaining potential drop lies across a large depletion region in the cathode adjacent ohmic region. Many experimentalists have speculated that this region should actually accumulate. In fact, at very low biases, this region does show slight accumulation. As the bias is increased, however, this region (near the cathode contact) begins to deplete of carriers. This depletion arises from a combination of sources. First, there is a triangular quantum well formed between the cathode contact and the first barrier potential which results in some quantization of the carriers (as the quantized level is moved further from the Fermi level, depletion arises), which is similar to that found in the inversion layer in an MOS transistor.[19] Secondly, the initialization of this triangular well is caused by the contact properties itself. There is an inherent contact resistance in this device, which is a manifestation of the Landauer contact resistance.[95] Moreover, cathode-adjacent depletion is well known in nonlinear two-terminal devices such as the Gunn-effect diode.[335] Such depletion results from the need to balance the resistance of the active region with that of the cathode source. The use of a spacer layer can alleviate this depletion.

DBRTDs are often fabricated with undoped spacer layers at the interface between the barriers and the bulklike regions of uniform doping. The effects that primarily accrue from the use of the spacer layer is to create a resistive region in the device, which can serve as a matching layer to the contact source of carriers, while also accumulating carriers for injection through the barriers. The role of the spacer layer on the self-consistent potential in equilibrium is shown in Fig. 31. It may be seen that a slight upward shift of the barriers arises from the undoped regions, and this will cause a much smaller (or no) depletion. This is a consequence of the better match of the overall device resistance to the contact properties, and will concentrate the potential into the barrier region, leading to a better peak-to-valley ratio. A second difference that will arise from the barrier layer, is a reduction in impurity scattering in the region near the barriers, which causes a longer mean-free path, and better tunneling characteristics. In Fig. 32, the potential distribution across the device is shown, both with and without the spacer layers, for a bias of 0.4 V, which is near the valley of the current in Fig. 29. It is clear that the spacer layers have essentially eliminated the depletion in the cathode adjacent region. There is a larger fraction of the potential dropped across the barrier region, and actually a small depletion in the anode adjacent region. The latter arises from the extraction characteristics of the anode contact. The I-V curves for a device with the spacer layers (5 nm thick) are shown in Fig. 33. It may be seen, from comparing this figure with Fig. 29, that the peak-to-valley ratio has been increased, both by enhancing the peak current through a lower series resistance and a lower valley current through the enhanced barrier potentials, and the negative conductance increased as well. In both figures, however, there is a well-defined hysterisis, or ***bistability***.

Intrinsic bistability in the DBRTD is thought to arise from charge storage in the quantum well. During the upsweep of the voltage, the quantum well resonant state sweeps through the full conduction band states, so that as the minimum of the current is approached the well is full of carriers (staying there through a sequential tunneling process, as opposed to a complete resonant tunneling process). On the downsweep of voltage, however, the well is basically empty as it comes into alignment with the full conduction band states. The difference in charge in the quantum well between the upsweep and downsweep changes the self-consistent potential, and the current, to reflect this bistability. This interpretation can be supported just by simple circuit arguments and from the presence of the negative differential conductance.[122]

The transient behavior of the DBRTD is readily studied as well, either by just switching the voltage from one level to another[336] or by actually superimposing a small-signal a.c. signal.[337] In Fig. 34, the current transient is shown for switching of the voltage from 0.26 V (at the peak of the $I-V$ curve) to 0.45 V (in the valley region), for the device without the spacer layers, in order to simulate a large signal response. This choice of voltages will select the intrinsic properties derived from discharging the quantum well, which is thought to lead to the large overshoot of current. The rapid oscillations are caused when the internal potential adjusts to the changing distribution, which is a coupling of plasma-like oscillations of the charge density to the RC controlled oscillations of the barrier capacitances. The carriers themselves, through their inductive response, cause the large overshoot of current. The large-signal transient has been calculated for three values of the relaxation scattering rate τ. These correspond to mobilities of 1500, 3000, and 4200 cm^2/Vs. The magnitude of the conductance is shown in Fig. 35(a). The real and imaginary parts of the condivity are plotted in parts (b) and (c) of the figure. The use of several different mobilities, and scattering times, insures that the basic response frequency is not due to the conductivity roll-off at frequencies such that $\omega\tau > 1$. In all three cases, the magnitude of the conductivity peaks at about 1.5 THz, although the peak height is sensitive to the magnitude of the mobility. This peak arises from the large reactive contribution to the conductivity, as may be seen from Fig. 35(b), even though the real part of the conductivity passes from negative to positive at this point, as shown in Fig. 35(c). It should be noted that this is almost an order of magnitude below the bulk carrier plasma frequency at the doping level used, so that the oscillation leading to this peak is not simply a plasma oscillation in the bulk density. Nor is it related strongly to the oscillations of the two-dimensional plasma in the well at the wave vector corresponding to the well thickness, as this frequency actually lies above the bulk plasma frequency. Nor is it a simple RC effect, since the change in the mobility would be expected to change R, and hence the RC roll-off frequency. Rather, the dominant time in these figures seems to be strongly related to the time necessary to tunnel through the resonant structure, although the latter is a controversial subject. However, this time (as estimated from the tunneling of Gaussian wave packets through resonant structures[338]) does correlate well with the peak frequency in Fig. 35. However, Tsuchiya *et al.*[339] report studies which show that the switching time correlates well with changes in the effective mass of the material system used, and *not well with barrier thicknesses.* This latter does not support a tunneling time interpretation, and especially not one in which the tunneling is sequential. It should be also noted that the frequency of the peak is quite near to the Bloch frequency

$$\omega_B = \frac{eEd}{\hbar} = \frac{eV}{\hbar}, \tag{136}$$

where V is the voltage drop across one of the barriers (the voltage drop between the quantum well and the states on the other side of one of the barriers). However, there is very little work on the relation of the Bloch frequency, which is usually defined in a periodic structure, and the states of a single quantum well (although this is the basis of Bloch oscillations in single-electron tunneling structures[137]), and it is clear that more work needs to be done to clarify the limiting frequencies of the DBRTD.

14. The Role of Dissipation

The above simulations were carried out in the relaxation-time approximation (85) with a constant scattering time. This is certainly an oversimplification of the scattering processes, and it would have been better to utilize at least an energy-dependent relaxation time, although there is no equation with which the local energy can be determined in the direct solutions of (30). The literature contains many allusions to an inaccuracy of the relaxation-time approximation itself, in that it is often asserted that it is not charge conserving. In fact, it should not be on the local basis, but should be so on the global basis (i.e., the total charge in a device must remain such that the overall device is charge neutral). Indeed, if we required that

$$\int d^3\mathbf{p} \left(\frac{f(\mathbf{R},\mathbf{p}) - f_0(\mathbf{R},\mathbf{p})}{\tau(\mathbf{p})} \right) = 0 , \tag{137}$$

there would not even be a drift-diffusion model. We must recall that the latter is found by rearranging the Boltzmann transport equation (1), in the relaxation-time approximation and in steady-state, to obtain

$$f(\mathbf{R},\mathbf{p}) = f_0(\mathbf{R},\mathbf{p}) - \tau(\mathbf{p}) \left[\frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{R}} - (\nabla V) \cdot \frac{\partial f}{\partial \mathbf{p}} \right] , \tag{138}$$

and if (137) were valid, there would be no current flow generated from this equation. Rather, (138) just illustrates that the current, which is obtained by taking the first moment with respect to momentum of this equation, arises from a balance between the driving forces (contained within the square brackets) and the dissipation (represented by τ).

To go beyond the simple model of a constant relaxation time used above, one must extend (30) to include the actual scattering processes in their natural physical basis. Levinson[314] included the quantum mechanical terms for inelastic scattering in his derivation of the Wigner equation of motion from the density matrix. These inelastic processes were included only to the equivalent of first-order in time-dependent perturbation theory, but they were done in a manner that incorporated the *intra-collisional field effect* (essentially, the existence of a non-zero collision duration that allows for an interference between the inelastic process and an accelerating potential).[30,340] Inelastic processes have subsequently been included in the simulation of the DBRTD,[341,342] but to relatively low order in perturbation theory. Nevertheless, the best summary of the various scattering processes is given by Frensley in the appendix to his review on the role of boundary conditions.[329]

In Fig. 36, the effect of the inelastic processes on the self-consistent potential within a DBRTD is shown, by varying the strength of the inelastic interaction.[342] It may be seen from this figure, that when the scattering is weak, a depletion layer forms in the cathode adjacent region within the device, just as described above. As the scattering strength is increased, this depletion is reduced due to the higher resistance of the doped regions, and at high scattering strengths no depletion layer is found in this area. This reinforces the above interpretation that the depletion region forms in order to match the active area resistance to the contact properties, and represents a contact resistance. Raising the resistance in the active device area, either with a high rate of inelastic scattering or by the introduction of a lightly-doped buffer layer adjacent to the barriers, eliminates the need for this depletion-induced resistance.

While the exact evaluation and comparison of the simulation with experimental data requires the inclusion of the details of the inelastic scattering processes, the major features of the operation of the DBRTD are recovered quite well with the constant scattering rate model. There remains, however, room to study the role of the inelastic processes further, since they have so far only been included to relatively low order in perturbation theory. How they may need to have the basic Wigner function renormalized, as is necessary in general with the Green's functions discussed in the next section, is currently not well known.

The coulomb interaction among the carriers has been included by obtaining the Wigner function from the real-time nonequilibrium Green's functions.[343] Here it is found that the scattering function contains memory effects arising from the real and advanced Green's functions. Under some simplifying approximations, the electron-electron interaction can be made to look like the electron-phonon interaction, at least within the form of the memory functions. The carrier-carrier interaction was also considered, along with the electron-phonon interaction, for carriers in a quantum well by Tso and Horing.[344] The interaction between the electrons and holes is shown to give a transient negative electron mobility, as seen experimentally in laser excitation experiments.[345]

15. Other Devices

Tsuchiya *et al.*[346] have used the Wigner function to simulate the linear and nonlinear transport in an electron waveguide, in which elastic scattering by impurities was included in the simulation. With weak scattering, they find the standard quantized conductance of the waveguide, given by (4). As the strength of the impurity scattering is increased, the steps in the conductance are found to deteriorate and to eventually disappear. At large bias voltages, the nonlinear transport also causes the conductance to deviate from (4). These authors have also included a gate across the quantum wire, and studied the transient switching of this gate, in analogy to that discussed above for the DBRTD. In Fig 37, the Wigner function is shown for two different bias voltages on the gate (the sign has been changed from that of the authors to represent voltage and not energy). The quantum wire is in a two-dimensional electron gas and is assumed to be 30 nm wide, with a Fermi energy of 10 mV at 4 K. Clearly, under a positive voltage, the region under the gate is accumulated, while a negative voltage depletes the region under the gate. When the gate voltage is switched from zero to 50 mV, creating a potential well under the gate, oscillations are observed in the current, which are considered to be due to the resonance of the electron wave in the potential well formed under the gate. Isawa and Hatano[347] have also used the Wigner function to study a one-dimensional ballistic channel with a single point scattering represented by a δ-function potential.

Miyoshi *et al.*[348] have used the Wigner function to study the transport of holes in the two-dimensional gas formed at the interface of a InGaAs-InP avalanche photodiode. In the structure, thin quaternary layers were inserted at the interface to eliminate hole trapping. Grading was also used for this purpose, with the latter found to be more effective. Scattering was modeled by a simple relaxation-time approximation. These authors have used the Wigner function approach to model both electrons and holes in a separate-confinement heterostructure quantum well laser.[349] They find that electrons and holes are not equally injected into the quantum well, which means that the conventional gain model needs to be changed.

V. Modeling with the Green's Functions

The previous two sections dealt with modeling devices in ways that extend quite easily to the semi-classical world. The approaches, as pointed out earlier, involve only a single time variable and concentrate on quantum effects that arrive from *spatial* correlation. The Green's functions, however, include two time variables, and describe propagation/correlation between two spatial points at different times. Thus, temporal correlation processes are incorporated in a more fundamental manner. With this added complication, for far more equations are needed to describe the Green's functions, comes the benefit of a much more direct incorporation of dissipative processes. In Sec. 5b, the set of nonequilibrium Green's functions was introduced. These different functions varied with the ordering of the field operators and time events. While six different Green's functions were introduced (the retarded and advanced functions, the less-than and greater-than functions, and the time-ordered and anti-time-ordered functions), only four of these are considered to be independent in the general nonequilibrium situation, and only two are independent in the equilibrium system. The difference between the nonequilibrium system and the equilibrium system lies in the added requirement in the former case to return the integration path to the initial time, as indicated in Fig. 8, if we are to properly retain normalization of the Green's functions. This normalization is necessary if we are to be able to cancel disconnected Feynman diagrams in a perturbation expansion.[204,213]

Keldysh introduced a general method of treating the set of Green's functions with a single matrix Green's function.[200] The Keldysh matrix may be written as

$$\mathrm{G} = \begin{bmatrix} G_t & G^> \\ G^< & -G_{\bar{t}} \end{bmatrix} . \tag{139}$$

We may think about this matrix form in the following way. The rows and columns of the matrix correspond to the time arguments of the Green's functions. While it is not correct to identify the rows and columns with the portions of the trajectory in Fig. 8, the "11" position is the Green's function that one would get if both time variables were on the bottom real time path (forward in time), and the "22" position is the Green's function one would get if both time variables were on the upper real time path (anti-time directed). The other two arise for the cases of contour ordering if the two time variables are on opposite parts of the contour, with the contour extension to the thermal equilibrium point being ignored. The extension to a 3 × 3 matrix to include the tail in Fig. 8 has been carried out by Wagner.[203] Actually, the Keldysh form rearranges these Green's functions slightly to obtain a zero in one corner of the matrix. If we subtract G_t from each term in the matrix, and then multiply through by a minus sign, the new Keldysh Green's function is just

$$\mathrm{G}_2 = \begin{bmatrix} 0 & G_a \\ G_r & G_K \end{bmatrix} , \tag{140}$$

where the Keldysh function is

$$G_K = G^> + G^< \,, \tag{141}$$

and the relations (40) and (41) have been used. Interchanging the rows of (140) leads to another commonly used form[350]

$$\mathbf{G}_3 = \begin{bmatrix} G_r & G_K \\ 0 & G_a \end{bmatrix} . \tag{142}$$

We can now develop the equations-of-motion for the non-interacting forms of these Green's functions; e.g., the equations which the functions will satisfy in the absence of any applied potentials and perturbing interactions. For this, we assume that the individual field operators are based upon wave functions which satisfy the basic Schrödinger equation (6). This leads to the two matrix equations

$$\left(i\hbar\frac{\partial}{\partial t} - H_0(\mathbf{r}) - V(\mathbf{r})\right)\mathbf{G}_{3,0} = \hbar\mathbf{I} \,, \tag{143}$$

$$\left(-i\hbar\frac{\partial}{\partial t'} - H_0(\mathbf{r}') - V(\mathbf{r}')\right)\mathbf{G}_{3,0} = \hbar\mathbf{I} \,, \tag{144}$$

where I is the unit matrix (unity on the diagonal and zeroes off the diagonal). The "0" subscript has been added to indicate that this is the non-interacting form (there is no inter-particle or particle-lattice interaction). In the next few sections, we will show how this is expanded to include the electron-phonon and other interactions, and how it is applied to study some device structures.

16. Homogeneous, Low-Field Systems

Transport, as has been stated earlier, arises as a balance between the driving forces and the dissipative forces. To achieve a description of transport with the Green's functions, it is necessary to now add some interaction terms to the Hamiltonian. The interaction term is treated as a perturbation in the usual case, along the lines of the S-matrix expansion of (35). These terms are usually expressable in terms of Feynman diagrams with the use of Wick's theorem.[213] In the present context, where it is desired to treat dissipation through the electron-phonon interaction, this procedure works relatively well, and has been almost universally used. The assumption is that projecting the time axes back to the initial time allows the use of the pseudo-equilibrium to justify the Wick's theorem expansion. The various parts of the diagrams may be regrouped then into terms which represent the Green's function itself and terms which represent the dissipative interaction with the lattice, which are referred to as the *self-energy*. The self-energy may also be expressed as a matrix Σ just as is the Green's function G itself. We will not treat the actual Feynman expansion here, as its form is available in numerous textbooks and review articles. Following this procedure, it is now possible to write the equation of motion for the full Green's function as

$$\left(i\hbar\frac{\partial}{\partial t} - H_0(\mathbf{r}) - V(\mathbf{r})\right)\mathbf{G} = \hbar\mathbf{I} + \Sigma\mathbf{G} \ , \tag{145}$$

$$\left(-i\hbar\frac{\partial}{\partial t'} - H_0(\mathbf{r}') - V(\mathbf{r}')\right)\mathbf{G} = \hbar\mathbf{I} + \mathbf{G}\Sigma \ , \tag{146}$$

where, in general, the Green's function matrix and the self-energy matrix are of the form (142). It should be noted that the reduced notation

$$\Sigma\mathbf{G}(1,3) = \int d2\Sigma(1,2)\mathbf{G}(2,3) \tag{147}$$

involves an integration over the included internal variables, and the shorthand notation $1 = (\mathbf{r}_1, t_1)$ is used, so that the integration is over three spatial variables and a time variable. It is necessary to point out here that the total self-energy has been split into two parts: (i) a single-site part that arises from the external potential $V(\mathbf{r})$ which does not require the integral of (147) since it occurs at a single point in space, and (ii) the nonlocal self-energy arising from the interaction of the electrons (or holes) with the phonons of the lattice or with other electrons (or holes). The normal Fermi-golden rule assumes that the interaction takes place instantaneously and at a single point in space, which are assumptions that reduce the nonlocal self-energy to a single-site representation of the self-energy. This approach has been followed by some, who purport to follow more general quantum transport,[351] but reduces the results to little more than the Boltzmann equation treatment.[228] A somewhat similar approach has been used by Datta and co-workers,[352] in which the self-energy for the scattering process is represented in the general form

$$\Sigma^{>}(\mathbf{r},\mathbf{r}';\omega) = -\frac{i\hbar}{\tau^{>}(\mathbf{r},\omega)}\delta(\mathbf{r}-\mathbf{r}') \ , \ \Sigma^{<}(\mathbf{r},\mathbf{r}';\omega) = \frac{i\hbar}{\tau^{<}(\mathbf{r},\omega)}\delta(\mathbf{r}-\mathbf{r}') \ , \tag{148}$$

where ω is the Fourier-transform variable corresponding to the difference in time $t-t'$. A similar description is used for the retarded self-energy, and this has been applied to treat especially impurity scattering in mesoscopic devices. Here, however, we will follow a more general treatment for the electron(or hole)-phonon interaction.

A. The Retarded Function

In low fields, the general approach is to seek the quantum transport equivalent of the Boltzmann equation. The traditional Boltzmann equation is expressed in terms of the distribution function $f(\mathbf{R},\mathbf{p},t)$. In previous sections, it was necessary to transform the density matrix or the Wigner distribution to achieve this quantity. Here, we want to describe the transport equation for low fields in a homogeneous system. This is, in essence, a linear approximation. In the process, we will introduce a phase-space distribution, along with the Wigner coordinates (the center of mass coordinates). The approach we follow is that of Mahan and co-workers,[353,354] and treat only the low electric field case. It has been extended to the case of both electric and magnetic fields.[355,356]

We begin by separating out the single entry in the matrix equation above for the retarded Green's function, which leads to the pair of equations

$$\left[i\hbar\frac{\partial}{\partial t}-H_0(\mathbf{r})-V(\mathbf{r})\right]G_r(\mathbf{r},\mathbf{r}')=\delta(\mathbf{r}-\mathbf{r}')+\int d^4\mathbf{x}''\Sigma_r(\mathbf{r},\mathbf{r}'')G_r(\mathbf{r}'',\mathbf{r}')\ ,$$

$$\left[-i\hbar\frac{\partial}{\partial t'}-H_0(\mathbf{r}')-V(\mathbf{r}')\right]G_r(\mathbf{r},\mathbf{r}')=\delta(\mathbf{r}-\mathbf{r}')+\int d^4\mathbf{x}''G_r(\mathbf{r},\mathbf{r}'')\Sigma_r(\mathbf{r}'',\mathbf{r}')\ . \tag{149}$$

These are now two equations that specify the retarded Green's function, and both must be satisfied. It is convenient at this point to introduce the change of variables (26), with the equivalent set for the time axes (T is the average time and τ is the difference time in this approach). Equations (149) now become, after adding and subtracting the two transformed equations

$$\left(i\hbar\frac{\partial}{\partial\tau}+\frac{\hbar^2}{8m}\nabla^2+\frac{\hbar^2}{2m}\frac{\partial^2}{\partial\mathbf{s}^2}+e\mathbf{E}\cdot\mathbf{R}\right)G_r(\mathbf{s},\mathbf{R},T,\tau)=\delta(\tau)\delta(\mathbf{s})$$

$$+\frac{1}{2}\int d^4\mathbf{x}''(\Sigma_rG_r+G_r\Sigma_r)\ ,$$

$$\left(i\hbar\frac{\partial}{\partial T}+\frac{\hbar^2}{m}\nabla\cdot\frac{\partial}{\partial \mathbf{s}}+e\mathbf{E}\cdot\mathbf{s}\right)G_r(\mathbf{s},\mathbf{R},T,\tau)=\int d^4\mathbf{x}''(\Sigma_r G_r-G_r\Sigma_r)\,. \quad (150)$$

The functions inside the integrals on the right-hand sides of these two equations have not yet been transformed to the new coordinates, as this is a complicated process with which we will deal below.

The two equations in (150) describe the relative motion about the center-of-mass motion and the latter motion itself. To proceed, it is now useful to Fourier transform the relative motion coordinates in order to achieve $G_r(\mathbf{k},\mathbf{R},T,\Omega)$. Then, the above equations become

$$\left(\hbar\Omega-E_k+\frac{\hbar^2}{8m}\nabla^2+e\mathbf{F}\cdot\mathbf{R}\right)G_r(\mathbf{k},\mathbf{R},T,\Omega)=1$$

$$+\frac{1}{2}\int d^3\mathbf{s}\,d\tau\, e^{i\mathbf{k}\cdot\mathbf{s}-i\Omega\tau}\int d^4\mathbf{x}''(\Sigma_r G_r+G_r\Sigma_r)\,,$$

$$i\hbar\left(\frac{\partial}{\partial T}+\mathbf{v}_k\cdot\nabla+\frac{e\mathbf{E}}{\hbar}\cdot\frac{\partial}{\partial\mathbf{k}}\right)G_r(\mathbf{k},\mathbf{R},T,\Omega)=\int d^3s\,d\tau\, e^{i\mathbf{k}\cdot\mathbf{s}-i\Omega\tau}$$

$$\times\int d^4\mathbf{x}''(\Sigma_r G_r-G_r\Sigma_r)\,. \quad (151)$$

The second of these equations has the same streaming terms on the left-hand side as the Boltzmann equation (1); however, the first equation has some problems which arise from the streaming terms in the large parentheses (the last term). This term gives rise to a driving force that has an unusual position dependence in the otherwise homogeneous system. This leads to a size dependence which is not at all in keeping with the physics of the structure. On the other hand, if this is combined with the "frequency" it is apparent that these two terms together represent the gauge variation of the energy and potential. This suggests a change of frequency (and spatial) variable as

$$\Omega+\frac{e\mathbf{F}\cdot\mathbf{R}}{\hbar}\rightarrow\omega\,,\ \nabla\rightarrow\nabla+\frac{e\mathbf{F}}{\hbar}\frac{\partial}{\partial\omega}\,. \quad (152)$$

With these changes, (151) become

$$\left[\hbar\omega-E_k+\frac{\hbar^2}{8m}\left(\nabla+\frac{e\mathbf{F}}{\hbar}\frac{\partial}{\partial\omega}\right)^2\right]G_r(\mathbf{k},\mathbf{R},T,\omega)=1$$

$$+\frac{1}{2}\int d^3\mathbf{s}d\tau e^{i\mathbf{k}\cdot\mathbf{s}-i\Omega\tau}\int d^4\mathbf{x}''(\Sigma_r G_r + G_r\Sigma_r)\,,$$

$$i\hbar\left[\frac{\partial}{\partial T}+\mathbf{v}_k\cdot\nabla+\frac{e\mathbf{E}}{\hbar}\cdot\left(\frac{\partial}{\partial\mathbf{k}}+\mathbf{v}_k\frac{\partial}{\partial\omega}\right)\right]G_r(\mathbf{k},\mathbf{R},T,\omega)=\int d^3\mathbf{s}d\tau e^{i\mathbf{k}\cdot\mathbf{s}-i\Omega\tau}$$
$$\times\int d^4\mathbf{x}''(\Sigma_r G_r - G_r\Sigma_r)\,. \tag{153}$$

The large omega in the integral stills needs to be transformed for these equations. It is clear that this variation in the exponential in the transformed collision terms leads to phase interference events for distances small compared to the inelastic mean free path.

The second of equations (153) contains a term that is linear in the field, but it is important to remember that the major factor leading to nonlinear transport is from the distribution function itself and not from nonlinear terms in the field. In fact this equation is exact to higher orders in the field. The scattering terms can be quite complicated, and some form of simplification is necessary to evaluate them to any great degree. Since the treatment in this sub-section is supposedly for small electric fields in an otherwise homogeneous medium, it is possible to use a gradient expansion.[201]

To proceed, the functions are expanded about the center-of-mass coordinates, which assumes that both of the latter are much larger than the relative coordinates. This expansion limits the applicability of the results to systems in which the potentials are slowly varying in both space and time. This is hardly the case in mesoscopic semiconductor devices, but is a good approximation for linear transport in a homogeneous semiconductor. In order to illustrate this procedure, only the first term in the collision integral on the right-hand side of (153) is considered. This integral is

$$I=\int d^3\mathbf{s}e^{i\mathbf{k}\cdot\mathbf{s}}\int d\tau e^{i\Omega\tau}\int d^4\mathbf{x}''\Sigma_r(\mathbf{x},\mathbf{x}'')G_r(\mathbf{x}'',\mathbf{x}')\,. \tag{154}$$

First, the center-of-mass coordinates $\mathbf{y}=\mathbf{x}-\mathbf{x}'$ are introduced so that the integral is now a function of $\Sigma_r(\mathbf{y},\mathbf{x}-\frac{\mathbf{y}}{2})G_r(\mathbf{x}-\mathbf{x}'-\mathbf{y},\frac{\mathbf{x}+\mathbf{x}'-\mathbf{y}}{2})$. Then, the coordinates are again transformed to $\mathbf{s}=\mathbf{x}-\mathbf{x}'$, $\mathbf{R}=\frac{\mathbf{x}+\mathbf{x}'}{2}$. Finally, the variable change $\mathbf{w}=\mathbf{s}-\mathbf{y}$ is made, and (154) becomes

$$I=\int\int d^3\mathbf{w}d^3\mathbf{y}e^{i\mathbf{k}\cdot(\mathbf{w}+\mathbf{y})}\int\int d\tau d\tau' e^{-i\Omega(\tau+\tau')}$$

$$\times\ \Sigma_r(\mathbf{y},\mathbf{R}+\frac{\mathbf{w}}{2},\tau',T+\frac{\tau}{2})G_r(\mathbf{w},\mathbf{R}-\frac{\mathbf{y}}{2},\tau,T-\frac{\tau'}{2})\,. \tag{155}$$

The center-of-mass variables all appear with offsets, so that these may be separated by expanding the functions in a Taylor series, and the Fourier transforms then taken so that the scattering function now becomes

$$I=\Sigma_r(\mathbf{k},\mathbf{R},\Omega,T)G_r(\mathbf{k},\mathbf{R},\Omega,T)+\frac{i}{2}\{\Sigma_r,G_r\}+...,\tag{156}$$

where the *bracket* operator is

$$\{\Sigma_r,G_r\}\equiv\frac{\partial\Sigma_r}{\partial\mathbf{R}}\cdot\frac{\partial G_r}{\partial\mathbf{k}}-\frac{\partial\Sigma_r}{\partial\mathbf{k}}\cdot\frac{\partial G_r}{\partial\mathbf{R}}+\frac{\partial\Sigma_r}{\partial\Omega}\cdot\frac{\partial G_r}{\partial T}-\frac{\partial\Sigma_r}{\partial T}\cdot\frac{\partial G_r}{\partial\Omega}\,.\tag{157}$$

This latter term has a natural symmetry that is also found in the Poisson brackets of classical mechanics, and the short-hand notation of the brackets is often used for this purpose. The frequency derivatives are with respect to the unshifted frequency at this point, so that we must still correct for the change in frequency. This latter operation leads to

$$\begin{aligned}\{\Sigma_r,G_r\}=&\frac{\partial\Sigma_r}{\partial\mathbf{R}}\cdot\frac{\partial G_r}{\partial\mathbf{k}}-\frac{\partial\Sigma_r}{\partial\mathbf{k}}\cdot\frac{\partial G_r}{\partial\mathbf{R}}+\frac{\partial\Sigma_r}{\partial\omega}\cdot\frac{\partial G_r}{\partial T}-\frac{\partial\Sigma_r}{\partial T}\cdot\frac{\partial G_r}{\partial\omega}\\&+\frac{e\mathbf{F}}{\hbar}\left(\frac{\partial\Sigma_r}{\partial\omega}\cdot\frac{\partial G_r}{\partial\mathbf{k}}-\frac{\partial\Sigma_r}{\partial\mathbf{k}}\cdot\frac{\partial G_r}{\partial\omega}\right)\,.\end{aligned}\tag{158}$$

Already the first term in the gradient expansion has produced a term that is linear in the electric field. If higher-order terms are retained, they lead to higher-order terms in the electric field. Since these terms have been neglected, the resulting expansion is valid only to linear terms in the field.

With these results, it is now possible to write down a relatively closed, but still approximate, set of equations for the retarded Green's function. Combining the above results, these equations become

$$\left[\hbar\omega-E_k+\frac{\hbar^2}{8m}\left(\nabla+\frac{e\mathbf{F}}{\hbar}\frac{\partial}{\partial\omega}\right)^2-\Sigma_r\right]G_r=1\,,$$

$$\left\{\frac{\partial}{\partial T}+\mathbf{v}_k\cdot\nabla+\frac{e\mathbf{F}}{\hbar}\cdot\left[\left(1-\frac{\partial\Sigma_r}{\hbar\partial\omega}\right)\frac{\partial}{\partial\mathbf{k}}+\left(\mathbf{v}_k+\frac{\partial\Sigma_r}{\hbar\partial\mathbf{k}}\right)\frac{\partial}{\partial\omega}\right]\right\}G_r=-\frac{i}{\hbar}\{\Sigma_r,G_r\}\,,\tag{159}$$

where the bracket on the right-hand side of (159) now refers only to the first term on the right-hand side of (158).

These equations simplify considerably for near equilibrium systems (which are the ones of interest) that are spatially homogeneous and in steady-state, so that the derivatives with respect to $\mathbf{R}$ and T vanish. Then, the above equations can be written as

$$[\hbar\omega - E_k - \Sigma_r]\, G_r = 1 \;,$$

$$ie\mathbf{F}\cdot\left[\left(1 - \frac{\partial\Sigma_r}{\hbar\partial\omega}\right)\frac{\partial}{\partial\mathbf{k}} + \left(\mathrm{v}_k + \frac{\partial\Sigma_r}{\hbar\partial\mathbf{k}}\right)\frac{\partial}{\partial\omega}\right] G_r = 0 \,. \tag{160}$$

The first equation can be solved quite simply to yield a result that will also satisfy the second equation for sufficiently small electric fields $\mathbf{F}$. The retarded Green' function is then finally found to be (retaining only terms linear in the field)

$$G_r(\mathbf{k},\mathbf{R},\omega,T) = \frac{1}{\hbar\omega - E_k - \Sigma_r(\mathbf{k},\mathbf{R},\omega,T)} + O(F^2) \,. \tag{161}$$

The retarded Green's function in this equation has no first-order terms in the electric field. Thus, its *form* is unchanged from that of the equilibrium function, although of course the self-energy may have undergone some changes. It may also be expected that the *spectral density* $A(\mathbf{k},\mathbf{R},\omega,T) = -2Im\{G_r(\mathbf{k},\mathbf{R},\omega,T)\}$ will also have the equilibrium form. The equilibrium form of this function is a Lorentzian, which arises from normal collisional-broadening. These important results will have a major effect on achieving a quantum Boltzmann equation below, after finding the "less-than" function. Before proceeding, however, it should be noted that if the quadratic terms in the field are retained in the retarded Green's function, then the inverse transform into the difference coordinates will contain Airy functions. This observation provides a guide to a powerful approach to the high-field terms in the next section.

B. The "Less-Than" Function

The next step is to repeat the above procedure for the "less-than" function $G^<$. It is this latter function that is related to the distribution function that we normally incorporate in the Boltzmann equation. As before, we will encounter two equations which arise from the two matrix equations (145) and (146). We can write these two equations as

$$\left[\hbar\omega - E_k + \frac{\hbar^2}{8m}\left(\nabla + \frac{e\mathbf{F}}{\hbar}\frac{\partial}{\partial\omega}\right)^2\right] G^< = \frac{1}{2} I^+ ,$$

$$i\hbar\left\{\frac{\partial}{\partial T} + \mathbf{v}_k\cdot\nabla + \frac{e\mathbf{F}}{\hbar}\cdot\left[\frac{\partial}{\partial\mathbf{k}} + \mathbf{v}_k\frac{\partial}{\partial\omega}\right]\right\} G^< = I^- , \tag{162}$$

where the scattering functions are given by

$$I^{\pm} = \int d^3\mathbf{s}\, d\tau\, e^{i[\mathbf{k}\cdot\mathbf{s} - (\omega - e\mathbf{F}\cdot\mathbf{R}/\hbar)\tau]} \int d^4\mathbf{x}''[\Sigma_r G^< + \Sigma^< G_a \pm G_r\Sigma^< \pm G^<\Sigma_a] . \tag{163}$$

In the absence of any collisions, the right-hand sides vanish, of course. The left-hand side of the second equation has the same form as the Boltzmann equation found earlier (except for the second term in the square brackets). The derivative with respect to the frequency suggests a constraint that can be placed on the connection between the classical and the quantum forms of the Boltzmann equation. However, if we properly connect the Wigner distribution function with the classical distribution function, then using (43) to form this function from the above equation will clearly illustrate that the frequency derivative will vanish from the classical limit.

The gradient expansion can now be used on the terms $I^{\pm}$ in the same manner as previously. Our interest will remain focused on the homogeneous, steady-state situation in semiconductors,where derivatives with respect to $\mathbf{R}$ and T may be neglected. This also means that the gradient expansion terms can be evaluated at the center-of-mass coordinates with the relative coordinates appearing in their Fourier-transformed form. Then, equations (162) become

$$\left[\hbar\omega - E_k - S + \frac{e\mathbf{F}}{\hbar}\left(\frac{\partial\Gamma}{\partial\mathbf{k}}\frac{\partial}{\partial\omega} - \frac{\partial\Gamma}{\partial\omega}\frac{\partial}{\partial\mathbf{k}}\right)\right] G^< = W\Sigma^< - \frac{e\mathbf{F}}{4\hbar}\cdot\mathbf{J}(A) ,$$

$$e\mathbf{F}\cdot\left[\left(1 - \frac{\partial S}{\partial\omega}\right)\frac{\partial}{\partial\mathbf{k}} + \left(\mathbf{v}_k + \frac{\partial S}{\hbar\partial\mathbf{k}}\right)\frac{\partial}{\partial\omega}\right] G^< + 2\Gamma G^< = A\Sigma^< + \frac{e\mathbf{F}}{\hbar}\cdot\mathbf{J}(W) , \tag{164}$$

where the following definitions have been introduced

$$G_r = G_a^\dagger = W - \frac{iA}{2} \; , \; \Sigma_r = \Sigma_a^\dagger = S - i\Gamma \; ,$$

$$\mathbf{J}(u) = \frac{\partial \Sigma^<}{\partial \omega}\frac{\partial u}{\partial \mathbf{k}} - \frac{\partial \Sigma^<}{\partial \mathbf{k}}\frac{\partial u}{\partial \omega} \; . \tag{165}$$

We recognize the spectral density discussed above in these definitions. Similarly, Γ is related to the lifetime of the states involved.

It is now possible to make a further simplification for the near-equilibrium situation. Since the field multiplies the two bracketed terms, the Green's function within those terms may be replaced by the field-independent form since we are only seeking the linear response terms. In equilibrium, it is possible to separate the less-than function with the ansatz $G^< = iP(\omega)A(\mathbf{k},\omega)$, where $P(\omega)$ is the Fermi-Dirac distribution function. Each term multiplying the electric field contains either a factor of $P(\omega)$ or the derivative of this quantity with respect to frequency, since both $G^<$ and $\Sigma^<$ are proportional to $P(\omega)$. In the absence of the field, $G^<$ is just the equilibrium form, and the two collision terms $2\Gamma G^<$ and $A\Sigma^<$ must balance each other (which is just another statement of detailed balance). Thus, it is the deviation of the distribution from the Fermi-Dirac form which leads to the transport. Thus, those terms involving the derivative of $P(\omega)$ that must be of importance, and indeed the other terms will drop out. Thus, the second equation of (164) becomes

$$ie\mathbf{F}\cdot\left[\left(\mathbf{v}_k + \frac{\partial S}{\hbar\partial \mathbf{k}}\right)\Gamma + (\hbar\omega - E_k - S)\frac{\partial \Gamma}{\hbar\partial \mathbf{k}}\right]A^2(\mathbf{k},\omega)\frac{\partial P(\omega)}{\partial \omega} = A\Sigma^< - 2\Gamma G^< \; . \tag{166}$$

This result is the steady-state homogeneous form of the quantum Boltzmann equation. The right-hand side demonstrates a gain-loss mechanism which is somewhat different from that normally seen, but it may easily be cast into the normal form. To linear terms in the field, this equation is exact and its derivation by Mahan and co-workers represents a major step forward in the use of the real-time Green's functions.[353]

It may be noted that the coefficient of the second term in the square brackets (left-hand side of the above equation) goes to zero at the peak of the spectral density. We can reasonably therefore neglect this second term, as will the velocity correction arising from the momentum derivative of S

in the first term. This latter term is a correction to the velocity that arises from the correction to the single-particle energies arising from the real-part of the self-energy, just as the second term is a dispersive velocity correction arising from lifetime effects. The neglect of these terms really means that we are ignoring any corrections to the velocity structure of the single-particle energy states due to the scattering processes, which is more or less akin to assuming that the self-energy shifts the bands in a rigid manner. Certainly this is appropriate to assume for the linear response case. Then, (166) can be rewritten as

$$A^2(\mathbf{k},\omega)e\mathbf{F}\cdot\mathbf{v}_k\Gamma\frac{\partial P(\omega)}{\partial\omega} = -i\left[A(\mathbf{k},\omega)\Sigma^< - 2\Gamma G^<\right] . \tag{167}$$

This may be rearranged to give

$$G^< = iA(\mathbf{k},\omega)\left[\frac{\Sigma^<}{2i\Gamma} - \frac{A(\mathbf{k},\omega)}{2}e\mathbf{F}\cdot\mathbf{v}_k\frac{\partial P(\omega)}{\partial\omega}\right] . \tag{168}$$

At this point, it is necessary to introduce a form for the self-energy. It is easy to show that it may be expressed in terms of the Green's function itself and a phonon Green's function (we treat here only the case of phonon scattering). Thus, $\Sigma^< = iD^<G^<$, but where the Green's function is evaluated at the shifted momentum and frequency according to the phonon frequency and momentum. That is, we can write the self-energy in the form

$$\Sigma^<(\mathbf{k},\omega) = \frac{2\pi}{\hbar}\sum_{\pm}\int\frac{d^3\mathbf{q}}{(2\pi)^3}\left|M(\mathbf{k},\mathbf{q})\right|^2 G^<(\mathbf{k}\pm\mathbf{q},\omega\pm\omega_q) , \tag{169}$$

where the term in the magnitude brackets is the matrix element and includes the phonon occupation factors, and the summation runs over emission and absorption processes. It has been assumed here that optical phonons, which are taken to be dispersionless, are the only phonons of interest. To proceed, we now will make the *ansatz* that (168) can be rewritten as

$$G^< = iA(\mathbf{k},\omega)\left[P(\omega) - \Lambda(\mathbf{k},\omega)e\mathbf{F}\cdot\mathbf{v}_k\frac{\partial P(\omega)}{\partial\omega}\right] , \tag{170}$$

where we now have to evaluate the function $\Lambda(\mathbf{k},\omega)$. If we use this definition in (169), we find

$$\Sigma^<(\mathbf{k},\omega) = iP(\omega)\Gamma - ie\mathbf{F}\cdot\mathbf{v}_k\frac{\partial P(\omega)}{\partial\omega}\frac{2\pi}{\hbar}\sum_{\pm}$$

$$\times \int \frac{d^3\mathbf{q}}{(2\pi)^3} |M(\mathbf{k},\mathbf{q})|^2 \frac{\mathbf{v}_{k\pm q}}{\mathbf{v}_k} A\Lambda(\mathbf{k} \pm \mathbf{q}, \omega \pm \omega_q) , \tag{171}$$

where (it should not be missed that a short-hand notation is used here, and that both A and Λ have the arguments listed for their product)

$$\Gamma = \frac{2\pi}{\hbar} \sum_{\pm} \int \frac{d^3\mathbf{q}}{(2\pi)^3} |M(\mathbf{k},\mathbf{q})|^2 A(\mathbf{k} \pm \mathbf{q}, \omega \pm \omega_q) . \tag{172}$$

This now gives us back our initial assertion on the Green's function (170) if the scattering kernal satisfies the integral equation

$$\Lambda(\mathbf{k},\omega) = \frac{A(\mathbf{k},\omega)}{2} + \frac{i}{2\Gamma(\mathbf{k},\omega)} \frac{2\pi}{\hbar} \sum_{\pm} \int \frac{d^3\mathbf{q}}{(2\pi)^3} |M(\mathbf{k},\mathbf{q})|^2 \frac{\mathbf{v}_{k\pm q}}{\mathbf{v}_k} A\Lambda(\mathbf{k}\pm\mathbf{q}, \omega\pm\omega_q) . \tag{173}$$

The quantity $\Lambda(\mathbf{k},\omega)$ can be referred to as a *vertex correction*, in that the integral equation is solved to find a modified scattering strength. For example, if we had assumed that there was also carrier-carrier scattering that modified the Green's functions, this would also appear in the scattering kernal Λ as a correction for the carrier response, which is one form of dynamic screening of the scattering process. It also has to be considered that if the phonon occupations are driven from their equilibrium forms, then the matrix element above is modified and a set of equations must be solved for the phonon Green's functions.

We note that the set of equations and the paths used to solve them is just as discribed in Sec. 2 above. We have first solved for the retarded Green's function in order to find the spectral density. We then solved for the "less-than" Green's function, which is related to the distribution function. The equation for this latter quantity was found to be very similar to the Boltzmann equation, at least in the linear response used here. A similar approach will be followed in nearly every case of quantum transport with the Green's functions. The introduction of the two-time formalism has given us the need to additionally find the spectral density. This was not necessary in the one-time formalism since there was an integration over the frequency (energy) domain.

17. Homogeneous, High-Field Systems

The development of a tractable quantum transport approach, through the use of the real-time Green's functions, which incorporates both the collisional broadening of the spectral density and the intracollisional field effect has proven difficult. (The latter is the interference between the driving terms in the field and the collisions that arises from the use of the nonequilibrium Green's functions in the self-energy.) Moreover, this task is further complicated by the need to deal with the length and time scales relevant to modern mesoscopic semiconductor devices. The general approach has followed that of the linear response approach of the last section. While the overall Green's function approach is rigorous in principle, most applications have been limited by the introduction of the center-of-mass approach and the gradient expansion. These two processes, especially the latter, tend to limit the results to low fields. One of the earliest to go beyond this was the work of Jauho and Wilkins,[357] in which high-field transport in a resonant-level impurity system of scattering was treated. In this work, the self-energies for electron-phonon as well as impurity scattering were formulated, although in the end the gradient expansion was invoked in order to achieve a result for the kinetic equation. Nevertheless, the power of the Green's function technique, when compared to other approaches, is still evident.[358]

In order to separate the "distribution function" part of the less-than Green's function from the spectral density, one has usually had to invoke an *ansatz*, such as that leading to (166) in which $G^< = iP(\omega)A(\mathbf{k},\omega)$ is assumed. The form of this ansatz has been the topic of some debate, primarily as there are several methods of approaching the separation. In fact, the ansatz is usually a method of recovering a single-time equivalent of the Boltzmann equation. The use of the gradient expansion is just one type of separation approach. A more rigorous approach, which is a zeroth-order approximation in the exansion of the correlation functions in terms of the collision duration time, has been developed by Lipavsky *et al.*[359,360] This form relates the Green's function to the density matrix as

$$G^<(t,t',\mathbf{p}_F) = -i\rho^<(t,\mathbf{p}_F)G_a(t,t',\mathbf{p}_F)\ , \tag{174}$$

where

$$\mathbf{p}_F = -i\hbar\nabla + e\int^t d\tau\mathbf{F}(\tau) \tag{175}$$

is the kinetic momentum operator. This approach was found to overcome a number of limitations and effects that arise from the normal ansatz used in the separation.[361] We will see below, however, that an approach using the Airy transforms brings out the separation automatically without having to make an unwarranted ansatz. Other approaches use expansions in the field,[362] or models for the spectral density.[363]

The other item mentioned above, that has been somewhat controversial, is the intra-collisional field effect. In general approaches to the quantum transport, a finite duration for the collision is found to occur, and the field can interact with the particle *while it is undergoing its collision.* Hence, the amount of energy and momentum gained or lost in the collision is modified by the effect of the field during the collision.[30,314] It was expected that this could make a significant effect on the transport properties.[12,32,340,364] In another approach, in which the collisions were treated as being completed, no intra-collisional field effect was observed.[365] Later work, however, showed that it would appear, but would be of such a size as to be negligible in normal completed collisions,[34] a result in keeping with numerical studies of the effect.[366] To date, there is no strong evidence that the intra-collisional field effect is a significant process, with the possible exception of impact ionization in wide-band gap semiconductors, in which very high fields are coupled with relatively slow processes.[367,368]

The collision duration itself is another problem, and one which still has to be sorted out. There is not much work applicable to this, but one approach using the Green's functions is due to Lipavsky *et al.*[369] These authors have also shown that different definitions of the instantaneous approximation for the self-energy will lead to different effects of the field on the collision, an effect that is negated if wave function, and therefore Green's function, renormalization is properly included.[370]

The use of the center-of-mass transformations introduces some non-physical variables into the description of transport. These, in fact, make the explicit assumption that the center-of-mass time $T = (t + t')/2$ has some inherent significance. This isn't the case, and nothing makes the point clearer than the need to modify the ansatz used to separate the distribution function from the less-than function as shown in (174). Consider the velocity autocorrelation function, which is a function of $t - t'$, where t' is the initial time of this function. The center-of-mass time doesn't enter into any consideration of physical quantities. The same consideration is true for the Green's functions of interest here. If a transformation is made to the center-of-mass coordinates, and homogeneity constraints (such as the assertion that the solutions are not functions of **R**) are applied to the space and time scales, this is fully equivalent to coarse-graining the Green's functions over time and space scales corresponding to the inelastic mean free path and inelastic mean free time. It is well-known, particularly in mesoscopic systems, that the local current is highly nonuniform, even in otherwise homogeneous systems, and that this leads to ***universal conductance fluctuations*** (d.c. fluctuations of the current with potential, rather than in time). This means that it will be impossible to utilize these center-of-mass formulations for such important applications as proper transient response and the calculation of fluctuation properties on a scale smaller than these coarse-graining times and lengths. Since many modern semiconductor devices have characteristic lengths smaller than, or comparable to, the inelastic mean free path (about 0.1 μm in Si at room temperature), these simplified approaches are doomed to failure in applications to these devices.

In fact, a constant electric field breaks the symmetry of the system, so that momentum along the field is no longer a good quantum number, at least on the scale of the inelastic mean free path. If the field is treated in the scalar potential gauge, it is the spatial translational symmetry that is broken, and this was seen to be the case in the approach of the last section where it was necessary to remove an artificial dependence upon position by a transformation of the frequency (energy) variable in (152). For this reason, many authors have placed the field in the vector potential, but this breaks the temporal translational symmetry, as evidenced from the results of (175). These facts have greatly complicated the search for high-field solutions for the Green's functions.

A. The Airy Function Retarded Green's Function

For the above reasons, a different approach is suggested for the treatment of high-field transport with the Green's functions if the proper space and time variations are to be retained. It was remarked above, in connection with the retarded Green's function, that retention of corrections due to the field introduced Airy functions in the inverse transformation to real space and time. Indeed, if the Schrödinger equation is solved in the presence of a uniform electric field, the basis functions that result are the Airy functions. The approach to be used here is to begin at the start by transforming the position variable along the electric field direction using the generalized Airy transform (nevertheless, it will be found that the treatment still is for the steady-state). This will allow us to diagonalize the unperturbed Green's function in the presence of the field alone, and to achieve a simpler form of the Dyson's equation.[371] Nevertheless, it is assumed that the system has translational symmetry in the coordinates *normal* to the field. The general Airy transform is defined by

$$F(\mathbf{k},s)=\int d^2\xi\int\frac{dz}{2\pi L}e^{i\mathbf{k}\cdot\xi}Ai\left(\frac{z-s}{L}\right)f(\xi,z)\,, \tag{176}$$

where ξ is the two-dimensional position vector in the (x,y)-plane and the field is assumed to be oriented along the z-axis. For the Green's function, of course, there are two Airy transformations on the two coordinates. In this transform space, a function that is diagonal in momentum k (assumed) and s variables is translationally invariant in the transverse plane but not in the z-direction.

The retarded Green's function, in the presence of the field, still satisfies (149). These two equations must be Airy transformed. In the absence of the self-energy, the solutions are the equilibrium Green's function with the field present, which may be shown to be

$$G^F_{r,0}(\mathbf{k},s,s',t,t')=-\frac{i}{\hbar}u_0(t-t')e^{-\frac{i}{\hbar}E_{\mathbf{k},s}(t-t')}\delta(s-s')\,, \tag{177}$$

where

$$E_{\mathbf{k},s}=\frac{\hbar^2k^2}{2m}+eFs \tag{178}$$

in parabolic bands. This form has the distinct advantage that the Fourier transformation of the difference variable in time leads to (for a temporally homogeneous system)

$$G^F_{r,0}(\mathbf{k},s,s',\omega)=\frac{\delta(s-s')}{\hbar\omega-E_{\mathbf{k},s}+i\eta}\,, \tag{179}$$

which has the generic form expected for the retarded Green's function. There is a problem in using this form of the retarded Green's function as it assumes that the field was turned on in the infinite past (as indicated by the convergence factor η), but this ignores the transient part of the integration path used for the real-time functions. However, this is consistent with the ignoring of the "tail" segment connecting to the thermal equilibrium. Nevertheless, the approach to be used here is only for the steady-state, and cannot handle the transient solution. The use of this equilibrium form allows us to write the equations (149) as integral equations, *which are the Dyson's equations.* In the Airy transformed form, the first of these equations becomes

$$G_r(\mathbf{k},s,s',\omega) = G^F_{r,0}(\mathbf{k},s,\omega)\left\{\delta(s-s') + \int ds"\Sigma_r(\mathbf{k},s,s",\omega)G_r(\mathbf{k},s",s',\omega)\right\} , \tag{180}$$

where the diagonality of the equilibrium function has been assumed in the prefactor of the integral on the right-hand side.

To proceed, it is now necessary to develop an approximation for the self-energy that appears in (180). The self-energy is given by the general Green's function expansion[204]

$$\Sigma_r(\mathbf{r},\mathbf{r}') = i[G_r(\mathbf{r},\mathbf{r}')D^>(\mathbf{r},\mathbf{r}') + G^<(\mathbf{r},\mathbf{r}')D_r(\mathbf{r},\mathbf{r}')] . \tag{181}$$

The operator ordering in the last term (note that there is no convolution integral in this operator product) is such that it vanishes as the denstiy goes to zero. As a result, for nondegenerate semiconductors, it is possible to ignore the second term. Moreover, in most semiconductors, the scattering is relatively weak and the self-energy can be calculated in the Born approximation—essentially the Green's function in the first term is replaced by its equilibrium, non-interacting form which is the free propagator. Since the retarded Green's function is still solved self-consistently within the Dyson's equation, collisional broadening will still appear in the final form, and the field's presence in the free propagator introduces the high-field effects within the scattering process. The phonon Green's function, for non-polar optical scattering, can be expressed as (for equilibrium phonons)

$$D^>_0 = -i\pi\sum_{\pm}|M_q|^2\,\delta(\omega \pm \omega_{\mathbf{q}}) , \tag{182}$$

and the matrix element includes the phonon occupation functions. The sum runs over the emission and absorption processes. With these two approximations, the retarded self-energy can be written as (where the carrier wave vector is split into its field-directed part and transverse part)

$$\Sigma_r(\mathbf{k},\mathbf{k}',\omega) = \int \frac{d^3\mathbf{q}}{(2\pi)^3} \sum_{\pm} |M_q|^2 \, G^F_{r0}(\mathbf{w},\mathbf{w}',\omega \pm \omega_{\mathbf{q}}) \,,$$
$$\mathbf{w} = \mathbf{k} + k_z \mathbf{a}_z - \mathbf{q} \,, \; \mathbf{w}' = \mathbf{k}' + k_z' \mathbf{a}_z - \mathbf{q} \,. \tag{183}$$

To be useful, the momentum dependence in the z-direction will be transformed back to real space and then Airy transformed. In general, the matrix element for non-polar optical phonons is not momentum dependent, so there is no complication in the matrix element from these transformations. Finally, the self-energy is found to be[371]

$$\Sigma_r(\mathbf{k},s,\omega) = \frac{m}{\hbar^2}\sqrt{\frac{m\Theta}{2}} \sum_{\pm} |M_q|^2 \left[M(s,\omega \pm \omega_0) + iN(s,\omega \pm \omega_0)\right] \,, \tag{184}$$

where

$$M(s,\omega) = Ai'^2(y) + yAi^2(y) \,, \; N(s,\omega) = Ai'(y)Bi'(y) + yAi(y)Bi(y) \,,$$
$$y = \frac{eFs - \hbar\omega}{\Theta} \,, \; \Theta^3 = \frac{3(e\hbar F)^2}{2m} \,. \tag{185}$$

The real and imaginary parts of the self-energy are plotted in Fig. 38 for two different values of the electric field (and for parameters appropriate to Si at room temperature). The oscillatory behavior of the real part of the self-energy is quite interesting and indicates that the interaction of the field and the scattering process is creating an equivalent quasi-two-dimensional behavior in the carrier gas, which is reinforced by looking at the imaginary part, which is related to the density of states. The step-like structure reinforces this interpretation.

Since the self-energy is diagonal in the Airy variable s, the integral in Dyson's equation can be removed by assuming that the retarded function is also diagonal in this variable or by taking a delta-function dependence for the difference. This then leads to the final form for the retarded Green's function

$$G_r(\mathbf{k},s,\omega) = \frac{1}{\hbar\omega - E_{\mathbf{k},s} - \Sigma_r(\mathbf{k},s,\omega)} \,, \tag{186}$$

which is the equilibrium *form*, although the field is present in the transformed energy and in the self-energy. This result differs from that of the previous section. The spectral density is then twice the negative imaginary part of the retarded Green's function, and is

$$A(\mathbf{k},s,\omega)=\frac{-2Im\{\Sigma_r(\mathbf{k},s,\omega)\}}{[\hbar\omega-E_{\mathbf{k},s}-Re\{\Sigma_r(\mathbf{k},s,\omega)\}]^2+[Im\{\Sigma_r(\mathbf{k},s,\omega)\}]^2}, \quad (187)$$

which, again, has the equilibrium form, but differs by the explicit incorporation of a field dependence. This is shown in Fig. 39, and the shift and distortion due to the field is evident. Note, however, that the normalization is maintained and a single-time function, which would integrate over this spectral density, does not exhibit any large effect from the field distortion, as has been discussed above.

B. The Less-Than Function

With the approximations above, it has been possible to achieve a good representation of the spectral function, which is correct (within the approximations) to all orders in the electric field. Although it is constrained to the case of weak scattering, it demonstrates both the intra-collisional field effect and the collisional broadening effects. With this spectral function, it is now possible to calculate the "less-than" Green's function $G^<$. In the low-field case of the previous section, it was found that we had to solve an integral equation for a scattering function Λ. The latter represented aspects of the deviation from the Fermi-Dirac distribution function. Here, an integral equation will also result, but in this case it will lead to the quantum distribution function directly, without the need for an ansatz. However, this distribution function will not be the Wigner distribution function. Rather, the quantum distribution function will be a function of the energy alone, just as is found in the equilibrium case. To obtain the Wigner distribution function, we will need to couple the quantum distribution with the spectral density and integrate out the frequency (energy).

Our starting point, as in the previous section, is the matrix equations (145) and (146), which after introducing the Fourier transforms on the lateral position and the difference time, and the Airy transform on the dimension along the applied field, becomes

$$(\hbar\omega-E_{\mathbf{k},s})G^<(\mathbf{k},s,s',\omega)=\int ds"[\Sigma_r(\mathbf{k},s,s",\omega)G^<(\mathbf{k},s",s',\omega)$$

$$+\; G^{<}(\mathbf{k},s,s",\omega)\Sigma_a(\mathbf{k},s",s',\omega)]\;,$$

$$(\hbar\omega - E_{\mathbf{k},s"})G^{<}(\mathbf{k},s,s',\omega) = \int ds"[G_r(\mathbf{k},s,s",\omega)\Sigma^{<}(\mathbf{k},s",s',\omega)$$

$$+\; \Sigma^{<}(\mathbf{k},s,s",\omega)G_a(\mathbf{k},s",s',\omega)]\;. \tag{188}$$

At this point, it has already been established that the retarded, and advanced by their connection, Green's functions and self-energies are functions of a single Airy variable. This eliminates the integration through the use of a delta function on the right-hand side of the latter equations. We can then formulate the sum and differences of these two equations, which in the present case yield the same resulting equation. In essence this result means that these two equations are self-adjoint. The resulting equation is just

$$G^{<}(\mathbf{k},s,s',\omega) = \frac{G_r(\mathbf{k},s,\omega) - G_a(\mathbf{k},s',\omega)}{E_{\mathbf{k},s} - E_{\mathbf{k},s"} + \Sigma_a(\mathbf{k},s',\omega) - \Sigma_r(\mathbf{k},s,\omega)}\Sigma^{<}(\mathbf{k},s,s',\omega)$$

$$= G_r(\mathbf{k},s,\omega)G_a(\mathbf{k},s',\omega)\Sigma^{<}(\mathbf{k},s,s',\omega)\;. \tag{189}$$

We note that the retarded and advanced Green's functions have imaginary parts sharply peaked around the diagonal, in fact we have taken them to be diagonal. In the last line of (189), we can now assume that the less-than self-energy is also diagonal without any great loss of generality. With this approximation, we can now rewrite (189) as

$$G^{<}(\mathbf{k},s,\omega) = iA(\mathbf{k},s,\omega)f(\mathbf{k},s,\omega)\;, \tag{190}$$

since the first line contains the spectral density in the numerator of the fraction, and the quantum distribution function is obtained, without any ansatz, to be

$$f(\mathbf{k},s,\omega) = \frac{\Sigma^{<}(\mathbf{k},s,\omega)}{2Im\{\Sigma_r(\mathbf{k},s,\omega)\}}\;. \tag{191}$$

Although at this point it is not obvious that this function will be independent of the momentum, this will be found to be the case below, at least for non-polar optical phonon scattering.

The less-than self-energy function can be developed as easily for the case of the non-polar optical phonon. As previously, it will be assumed that this function is diagonal in the Airy variable, so that we can write it (as in the previous section) as

$$\Sigma^{<}(\mathrm{k},s,\omega)=\frac{1}{3^{1/6}}\sum_{\pm}|M_q|^2\int d^2\mathrm{q}_t\int ds' As(\frac{s-s'}{3^{1/3}})G^{<}(\mathrm{k}-\mathrm{q}_t,s',\omega\pm\omega_{\mathrm{q}}),$$

$$As(y)=\frac{1}{L}Ai(\frac{y}{L}),\ L=\left(\frac{\hbar^2}{2meF}\right)^{1/3},\tag{192}$$

which depends, of course, on the less-than Green's function. Now, we note an important fact about this latter equation. The matrix element for the non-polar optical phonon is independent of the phonon momentum vector, so that any change of variables in the integration over this momentum will not affect the final result for the self-energy. In fact, ***such a change of variables will integrate out all of the momentum dependence on the right-hand side of the equation.*** This means that the less-than self-energy is a function only of the Airy variable s and the frequency (energy) ω. A casual look at (184) shows that the retarded self-energy likewise is a function only of these variables. Thus, the quantum distribution function can only be a function of these two variables. With these thoughts, the integral equation for the quantum distribution function becomes

$$f(s,\omega)=\frac{1}{2(3)^{1/6}Im\{\Sigma_r(\mathrm{k},s,\omega)\}}\sum_{\pm}|M_q|^2$$

$$\times\int ds' As(\frac{s-s'}{3^{1/3}})N(s',\omega\pm\omega_{\mathrm{q}})f(s',\omega\pm\omega_{\mathrm{q}}),\tag{193}$$

where

$$N(s',\omega)=\int d^2\mathrm{q}_t A(\mathrm{q}_t,s',\omega)\tag{194}$$

is the effective density of states at the final energy of the scattering process. We note that the denominator of the prefactor plays the same role as Γ in the low-field equations. In Fig. 40, such a distribution function calculated for the same conditions as the earlier plots (Si at room temperature) is shown. It should be remarked that the integral equation for the distribution function has a certain characteristic form typical of that obtained from the Boltzmann equation as well. This integral equation can be retransformed into real space, and a Monte Carlo procedure developed for it.[245]

It is not obvious that the present formulation retains the desired property of gauge invariance. This topic has been considered in connection with the Airy transform and the normal Keldysh approach.[372] It has been shown that gauge invariance carries through to this approach in a relatively straightforward manner. Moreover, the resulting form for the Green's functions in the Airy approach can be transformed directly back into the more normal formalism in a manner that maintains the gauge invariance.[373] Finally, we note that the Airy Green's functions can be connected directly to the Kubo formula for the current, and the resulting conductivity can be expressed as[374]

$$\sigma(F) = \frac{\hbar}{2} \int \frac{d^2\mathbf{k}}{(2\pi)^2} \int d\Omega \Big[\left(1 - \frac{\partial Re\{\Sigma_r\}}{\hbar \partial \Omega} \right) +$$

$$+ \ (\hbar\Omega - E_{\mathbf{k}} - Re\{\Sigma_r\}) \frac{\partial \ln(Im\{\Sigma_r\})}{\hbar \partial \Omega} \Big] A^2(\mathbf{k}, \Omega) f(\Omega) \,, \qquad (195)$$

where the s variation has been removed from the kinetic energy and placed in the variable $\Omega = \omega - eFs/\hbar$.

18. Femtosecond Laser Excitation

The transport of photoexcited carriers is usually described with the Boltzmann transport equation, in which scattering processes, and the excitation process itself, are described via Fermi golden rule transition rates. However, today it is possible to excite a semiconductor with a laser pulse that is as short as 6 fs.[375] It is no longer a simple case of exciting an electron-hole pair with a photon from one of these lasers, but now one must begin to deal with the correlation that exists between this electron and hole on the short time scale (< 100 fs). Indeed, the polarization of the pair itself can now be measured with some degree of accuracy.[376–380] In general, the semi-classical description will break down on this time scale. First, an optical pulse creates carriers in quantum states which are definitely not semi-classical in that the initial carrier wave functions are super-positions of conduction and valence band states.[381] As long as this phase coherence between the electron and hole states—the interband polarization—remains important, the carriers are not characterizable as plane waves with characteristic momenta from either the conduction or the valence band. Hence, on the short time scale, this polarization must be taken explicitly into account if the system is to be described with any sort of accuracy. The second major point is that the short time scale of the optical pulse gives rise to a very broad Fourier spectrum of photons impinging upon the semiconductor. The broadening in excitation photon spectrum that comes from this short time interaction must be taken into account in a manner equivalently to the collisional broadening.

The approach to treating rapid excitation of electron-hole pairs in semiconductors is a problem that is inherently a natural for the Green's function technique, although this study is relatively recent in application. In fact, one will ultimately want to reduce the treatment to a single time variable, just as in the density matrix approach. However, the inherent correlations that must be treated make this a problem better treated by the Green's functions where the temporal correlations are a natural result of the basis. In fact, the problem has been treated to some degree both by the density matrix[382] as well as by the real-time Green's function technique.[251,383–385] On the short-time scales, the coherence introduced by the laser pulse cannot be neglected, so that the interband polarization must be treated as another kinetic variable. The resulting developement leads to the optical Bloch equation.[252,386–388]

The presence of the correlation, that is represented by the interband polarization, is quite significant. The shock of the intense laser pulse fully, and completely, separates the electron-hole gas from the thermal equilibrium background. While there has been some work on treating the development of the non-equilibrium system from the thermal distribution (along the tail of the contour of Fig. 8),[389,390] this approach is not possible in the current context. This means that it is generally not possible to utilize Wick's theorem to discard the normal-ordered products in the Green' function expansion of Feynmann diagrams. These products are part of the initial correlations that remain after the laser excitation, and must be treated by higher-order diagrams in the Green's function expansion.[391,392] In this section, we want to describe the manner in which the initial correlations, and the carrier-carrier interactions, fit into the development of the Green's function equation of motions for the two carrier types.

In order to achieve the goal of writing the Green's functions as products of two types of carriers, we expand the definition of the Keldysh matrix given by, for example, (142). We consider that each of the field operators can be either an electron or a hole operator (an electron in the valence band). Then it is possible to write the equivalent of (142) as

$$\hat{\mathbf{G}} = \begin{bmatrix} \mathbf{G}_{cc} & \mathbf{G}_{cv} \\ \mathbf{G}_{vc} & \mathbf{G}_{vv} \end{bmatrix} , \tag{196}$$

which is a resulting 8 × 8 matrix, where here c and v stand for conduction and valence band carriers. The end result will be 8 equations for Green's functions for various types of particles and their interactions. Of course, there are certain symmetry properties among the various Green's functions. In the laser excited situation, however, we are interested in the equivalent density matrices for the two carrier types and in the polarization (the cross-terms in the subscripts of the above equation). This means that the final set of equations will have a single time variable, although we will develop these equations from the two-time formalism. Various approaches have appeared in the literature,[251,383,393] but the approach we will follow is largely due to Kuhn.[394]

For the investigation of fast laser excitation of semiconductors, and the consequent coherent dynamics, we need to explicitly consider the coupling between the carriers and the light field and to include these terms in the system Hamiltonian. Here, we will take the equal time form of the Hamiltonian (replacing the field operators by simple second-quantized creation and annihilation operators)

$$\begin{aligned} H_0 = {} & \sum_{\mathbf{k}} E_k^e c_{\mathbf{k}}^+ c_{\mathbf{k}} + \sum_{\mathbf{k}} E_k^h d_{\mathbf{k}}^+ d_{\mathbf{k}} + \sum_{\mathbf{q}} \hbar\omega_{\mathbf{q}} b_{\mathbf{q}}^+ b_{\mathbf{q}} \\ & + \sum_{\mathbf{k}} [\mathbf{M}_{\mathbf{k}} \cdot \mathbf{F}^{(+)} c_{\mathbf{k}}^+ d_{-\mathbf{k}}^+ + \mathbf{M}_{\mathbf{k}}^* \cdot \mathbf{F}^{(-)} d_{-\mathbf{k}} c_{\mathbf{k}}] , \end{aligned} \tag{197}$$

where $c_{\mathbf{k}}^+$, $d_{\mathbf{k}}^+$, $b_{\mathbf{q}}^+$ ($c_{\mathbf{k}}$, $d_{\mathbf{k}}$, $b_{\mathbf{q}}$) denote the creation (annihilation) operators for the electrons, holes, and phonons, respectively, $E_k^e = E_G + \hbar^2 k^2/2m_e$ and $E_k^h = \hbar^2 k^2/2m_h$ are the electron and hole energies (measured from the valence band edge), $\mathbf{M}_{\mathbf{k}}$ is the dipole matrix element connecting the hole state with the electron state in optical absorption, and $\mathbf{F}^{(\pm)}$ is the Fourier coefficient of the field at positive and negative frequency by

$$\mathbf{F}^{(+)} = \mathbf{F}_0(t) e^{-i\omega_L t} , \ \mathbf{F}^{(-)} = \mathbf{F}^{(+)*} , \tag{198}$$

with ω_L the center frequency and $\mathbf{F}_0(t)$ the optical pulse shape. The basic variables now used to describe the system response to the optical fields are the distributions of the electrons and holes

$$f^e_{\mathbf{k}} = < c^+_{\mathbf{k}} c_{\mathbf{k}} > \,, \; f^h_{\mathbf{k}} = < d^+_{\mathbf{k}} d_{\mathbf{k}} > \,, \tag{199}$$

and the interband polarization

$$p_{\mathbf{k}} = < d_{-\mathbf{k}} c_{\mathbf{k}} > \,, \; p^*_{\mathbf{k}} = < c^+_{\mathbf{k}} d^+_{-\mathbf{k}} > \,. \tag{200}$$

It will be noted that these are equal time Green's functions, and therefore are actually mixed density matrices. It will further be noted that these four functions are the four blocks of the super-Green's matrix (196). The Hamiltonian (197) does not have any interactions among the various excitations (electrons, holes, and polarizations), and in the absence of these, the equations of motion become

$$\frac{df^e_{\mathbf{k}}}{dt} = \frac{df^h_{-\mathbf{k}}}{dt} = g^0_{\mathbf{k}}(t) = \frac{1}{i\hbar}[\mathbf{M}_{\mathbf{k}}\cdot\mathbf{F}_0(t)e^{-i\omega_L t}p^*_{\mathbf{k}} - \mathbf{M}^*_{\mathbf{k}}\cdot\mathbf{F}^*_0(t)e^{i\omega_L t}p_{\mathbf{k}}] \,, \tag{201}$$

$$\frac{dp_{\mathbf{k}}}{dt} = \frac{1}{i\hbar}[(E^e_k + E^h_{-k})p_{\mathbf{k}} + \mathbf{M}_{\mathbf{k}}\cdot\mathbf{F}_0(t)e^{-i\omega_L t}(1 - f^e_{\mathbf{k}} - f^h_{-\mathbf{k}})] \,. \tag{202}$$

The main difference between the coherent generation rate $g^0_{\mathbf{k}}(t)$ and the Fermi-golden rule of optical absorption is that in this approach the generation of carriers is a two-step process. The light wave first creates a polarization excitation in the semiconductor, and then this polarization is *ionized* into a separate electron and hole. The connection with the previous section is made by, for example,

$$G^<_{eh}(\mathbf{k},t,t') = i < d_{-\mathbf{k}}(t')c_{\mathbf{k}}(t) > \,, \tag{203}$$

and this strange state is achieved by noting that $d_{-\mathbf{k}}$ *creates* a hole in the state $-\mathbf{k}$, which in effect *annihilates* an electron in the valence band electron state $\mathbf{k}$. This change in notation is a result of a "rotation" of valence electron states into hole quasi-particle states. (In reading the various literature on this topic, one should be aware of the vagaries of notation used. The notation here is that of Kuhn[394] while the alternative notation in terms only of electron operators is used by e.g. Haug and co-workers.[383]) The rotation and renormalization in terms of quasi-particles is often referred to as a Bogoliubov transformation.

Dissipative processes can be added to (201) and (202) quite easily. The interaction with phonons is directly added to the equations for the electron and hole distribution functions through self-energy terms as described in the last section (the limitation to equal times simplifies the process immensely). In fact, in this formulation, the electron and hole losses to the phonons can almost be as easily treated by semi-classical scattering terms, and this is one approach that has been used.[252,388,394] On the other hand, the inclusion of a collision broadening of the appropriate spectral function requires the retention of the proper Green's functions expansions, and this has been pursued by Haug and co-workers,[383] among others. In Fig. 41, the energy distribution for the electrons is shown at different times for the Bloch equation model treated by Kuhn and co-workers, and is compared to semi-classical results (ignoring the polarization effects). The multiple peaks are phonon replicas of the initial excitation. No carrier-carrier interactions are included in the model. In Fig. 42, the average kinetic energy and the polarization are shown for comparison. A 100 fs pulse width is assumed in these calculations.

Carrier-carrier interactions can be added to the process by including an additional term in the Hamiltonian, which has the usual form (again, written in a single time formulation)

$$H_1^{cc} = \sum_{\mathbf{k},\mathbf{k}',\mathbf{q}} V_q [\frac{1}{2} c^+_{\mathbf{k}+\mathbf{q}} c^+_{\mathbf{k}'-\mathbf{q}} c_{\mathbf{k}'} c_{\mathbf{k}} + \frac{1}{2} d^+_{\mathbf{k}+\mathbf{q}} d^+_{\mathbf{k}'-\mathbf{q}} d_{\mathbf{k}'} d_{\mathbf{k}} - c^+_{\mathbf{k}+\mathbf{q}} d^+_{\mathbf{k}'-\mathbf{q}} d_{\mathbf{k}'} c_{\mathbf{k}}] , \tag{204}$$

where the three terms represent electron-electron, hole-hole, and electron-hole interactions. In addition, the interband exchange term has also been neglected (in which the electron and hole exchange momenta). Using the Heisenberg equations of motion for the various density matrices, we find the extra term for the time rate of change of the electron distribution function to be

$$\begin{aligned} \left.\frac{df^e_{\mathbf{k}}}{dt}\right| = \frac{1}{i\hbar} \sum_{\mathbf{k}',\mathbf{q}} V_{\mathbf{q}} [&< c^+_{\mathbf{k}} c^+_{\mathbf{k}'-\mathbf{q}} c_{\mathbf{k}'} c_{\mathbf{k}+\mathbf{q}} > - < c^+_{\mathbf{k}-\mathbf{q}} c^+_{\mathbf{k}'-\mathbf{q}} c_{\mathbf{k}'} c_{\mathbf{k}} > \\ &+ < c^+_{\mathbf{k}+\mathbf{q}} d^+_{\mathbf{k}'-\mathbf{q}} d_{\mathbf{k}'} c_{\mathbf{k}} > - < c^+_{\mathbf{k}} d^+_{\mathbf{k}'-\mathbf{q}} d_{\mathbf{k}'} c_{\mathbf{k}-\mathbf{q}} >] , \end{aligned} \tag{205}$$

and analogous equations are obtained for the hole distribution function and the polarization. If Wick's theorem is used to factorize the products of the four operators into separate distribution functions and polarizations, then energy shifts are obtained corresponding to the Hartree and Hartree-Fock (exchange) contributions. The polarization leads to a renormalization of the internal field that excites the carriers. There are, however, second-order contributions, which are the higher-order correlation functions.[391,392] These two-particle correlations, when treated in the lowest order, lead to terms (and the need for equations of motion for these terms)

$$< c_{\mathbf{k}}^{+} c_{\mathbf{k}'-\mathbf{q}}^{+} c_{\mathbf{k}'} c_{\mathbf{k}+\mathbf{q}} > + f_{\mathbf{k}}^{e} f_{\mathbf{k}+\mathbf{q}}^{e} \delta_{\mathbf{k},\mathbf{k}'} \,, \quad < c_{\mathbf{k}+\mathbf{q}}^{+} d_{\mathbf{k}'-\mathbf{q}}^{+} d_{\mathbf{k}'} c_{\mathbf{k}} > - p_{\mathbf{k}+\mathbf{q}}^{*} p_{\mathbf{k}} \delta_{\mathbf{k},-\mathbf{k}'} \,. \tag{206}$$

In general, these lowest-order terms now lead to the lifetimes of the electron, hole, and polarization states.[395] However, it is not at all clear that the higher-order terms can be ignored, and normally, the four-operator Green's function must be determined from a self-consistent integral equation, often called the Bethe-Salpeter equation. In fact, these terms are also important for dynamic screening of the carrier interactions, and the basic electron-electron term leads to what is also known as the polarization bubble (polarization here of the electron gas separately). These higher-order terms are also important for weak localization and universal conductance fluctuations at low temperatures.[14,15] To date, few of these more complicated approaches have been utilized in the femtosecond excitation world, and this area remains one in which many fruitful advances are still to be made.

In keeping with the general treatment of the scattering as semi-classical, it is also possible in numerical simulations of the femtosecond response to include the electron-electron interaction in real space as a molecular dynamics potential, and this has led to some understanding of the effect of the interacting gas on this time scale.[396] There have also been attempts to include the finite duration of the collision that would arise by retaining the full two-time Green's function approach to the femtosecond response.[397]

19. The Green-Kubo Formula

In Sec. 6 above, the Kubo formula was developed in linear response to the applied fields, as represented in the vector potential. It was noted there that the use of the Kubo formula was a significant change from the normal treatment of the dominant streaming terms of the Boltzmann equation, or of the equivalent quantum transport equation, to the relaxation and/or scattering terms. With the Kubo formula, one concentrates on the relaxation processes through the correlation functions that describe the transport. Here, we would like to close this loop once again and talk about how the real-time

Green's functions fit into the Kubo formula, with the combination termed the Green-Kubo formula. Now, if we note that the quantum mechanical current is described by

$$j = \frac{e\hbar}{2im}[\Psi^*(r)\frac{\partial\Psi(r)}{\partial r} - \frac{\partial\Psi^*(r)}{\partial r}\Psi(r)] \,, \tag{207}$$

then it is not too difficult to develop the Green's function form of this quantity to use in the Kubo formula (61). Using these ideas, it is possible to write the equivalent Kubo formula in terms of the Green's function for the carrier system as

$$< \mathbf{j}(\mathbf{r},t) >= -\frac{ie^2\hbar}{4m^2} \lim_{\mathbf{r}'\to\mathbf{r},t'\to t}(\nabla - \nabla')\cdot \int dt_s \int d^3\mathbf{s} \lim_{\mathbf{s}'\to\mathbf{s},t'_s\to t_s}(\nabla_{\mathbf{s}} - \nabla_{\mathbf{s}'})$$

$$\times [G_r(\mathbf{r},\mathbf{s}',t,t'_s)G^<(\mathbf{s},\mathbf{r}',t_s,t') + G < (\mathbf{r},\mathbf{s}',t,t'_s)G_a(\mathbf{s},\mathbf{r}',t_s,t')]\mathbf{A}(\mathbf{s},t_s) \,, \tag{208}$$

where we carry out the indicated integrations and limits, and $\mathbf{A}(\mathbf{r},t)$ is the vector potential

$$\mathbf{A}(\mathbf{r},t) = -\int^t dt'\mathbf{F}(\mathbf{r},t') \,. \tag{209}$$

It should be noted here that the displacement current has been ignored. Now, this can be Fourier transformed to give the a.c. conductance

$$\sigma(\mathbf{k},\omega) = -\frac{e^2\hbar}{2m^2\omega}\int\frac{d^3\mathbf{k}_2}{(2\pi)^3}\int\frac{d\omega'}{2\pi}(\mathbf{k}_2 + \frac{\mathbf{k}}{2})\cdot\mathbf{k}$$

$$\times \left[G_r(\mathbf{k}_2 + \frac{\mathbf{k}}{2},\omega')G^<(\mathbf{k}_2 - \frac{\mathbf{k}}{2},\omega' - \omega)\right.$$

$$+ \left. G^<(\mathbf{k}_2 + \frac{\mathbf{k}}{2},\omega')G_r(\mathbf{k}_2 - \frac{\mathbf{k}}{2},\omega' - \omega)\right] \,, \tag{210}$$

which, after some further changes of variables becomes

$$\sigma(\mathbf{k},\omega) = -\frac{e^2\hbar}{m^2\omega}\int\frac{d^3\mathbf{k}_1}{(2\pi)^3}\int\frac{d\omega'}{2\pi}\mathbf{k}_1\cdot\mathbf{k}$$

$$\times \; [G_r(\mathbf{k}_1,\omega')G^<(\mathbf{k}_1 - \mathbf{k},\omega' - \omega)$$

$$+ \; G^{<}(\mathbf{k}_1,\omega')G_r(\mathbf{k}_1-\mathbf{k},\omega'-\omega)] \; . \tag{211}$$

It should be noted that this conductivity does not diverge as the frequency tends to zero, due to the integral over the shifted frequency. This form has certain advantages, but it has not been utilized very much with the real-time nonequilibrium Green's functions.

It is now useful to rearrange the terms into those more normally found in the equilibrium, and low temperature forms of the Green's functions. For this, we make the ansatz

$$G^{<}(\mathbf{k},\omega) = if(\omega)A(\mathbf{k},\omega) = -f(\omega)[G_r(\mathbf{k},\omega) - G_a(\mathbf{k},\omega)] \, . \tag{212}$$

This ansatz was shown to be correct for the low-field Green's function and for the Airy transformed Green's functions in previous sections. It is also correct in equilibrium for zero-temperature and thermal Green's functions, but it is not in keeping with the normal high-field ansatze that have been used in the Kadanoff-Baym and Keldysh representations.[359] Nevertheless, we shall pursue this definition. Using this in (211) leads us to

$$\begin{aligned}\sigma(\mathbf{k},\omega) = \frac{e^2\hbar}{m^2\omega}\int\frac{d^3\mathbf{k}_1}{(2\pi)^3}\int\frac{d\omega'}{2\pi}\mathbf{k}_1{\cdot}\mathbf{k}\,\{&G_r(\mathbf{k}_1,\omega')G_r(\mathbf{k}_1-\mathbf{k},\omega'-\omega)f(\omega') \\ &- G_a(\mathbf{k}_1,\omega')G_a(\mathbf{k}_1-\mathbf{k},\omega'-\omega)f(\omega'-\omega) \\ &- G_r(\mathbf{k}_1,\omega')G_a(\mathbf{k}_1-\mathbf{k},\omega'-\omega)[f(\omega')-f(\omega'-\omega)]\} \; .\end{aligned} \tag{213}$$

The first two products of Green's functions will cancel one another. This can be seen by changing the frequency variables as $\omega'' = \omega' - \omega$, then using the fact that $\sigma(\mathbf{k},\omega) = \sigma^*(\mathbf{k},-\omega)$, and $G_r = G_a^*$. Thus, we are left just with the last term in the curly brackets. For low frequencies, the distribution function can be expanded about ω', so that

$$\sigma(\mathbf{k},\omega) = -\frac{e^2\hbar}{m^2}\int\frac{d^3\mathbf{k}_1}{(2\pi)^3}\int\frac{d\omega'}{2\pi}\mathbf{k}_1{\cdot}\mathbf{k}G_r(\mathbf{k}_1,\omega')G_a(\mathbf{k}_1-\mathbf{k},\omega'-\omega)\frac{\partial f(\omega')}{\partial\omega'} \, . \tag{214}$$

Finally, at low frequencies and for homogeneous material, we arrive at the form

$$\sigma(0) = -\frac{e^2\hbar}{m^2}\int\frac{d^3\mathbf{k}_1}{(2\pi)^3}\int\frac{d\omega'}{2\pi}\mathbf{k}_1{\cdot}\mathbf{k}\,|G_r(\mathbf{k}_1,\omega')|^2\,\frac{\partial f(\omega')}{\partial\omega'} \, . \tag{215}$$

In the case of low temperatures, the derivative of the distribution function is essentially a delta function at the Fermi energy. For this case, and with a change to the velocities in the arguments, one arrives at

$$\sigma(0) = \frac{e^2}{2\pi\hbar} \int \frac{d^3\mathbf{k}_1}{(2\pi)^3} \mathrm{v}_1 \cdot \mathrm{v} \left| G_r(\mathbf{k}_1, \omega_F) \right|^2 , \tag{216}$$

where $\omega_F = E_F/\hbar$. The sum over the momentum counts the number of states that contribute to the conductivity and results in the density at the Fermi energy. In mesoscopic systems, where only a single transverse state may contribute, the Landauer formula (4) (an additional factor of two arises from a summation over spin states) is recovered when it is recognized that $|G_r(\mathbf{k}_1, \omega_F)|^2$ is the transmission for the mode. Even if there is no transverse variation, the integration over the longitudinal component of the wave vector will produce the difference in the Fermi energies at the two ends of the samples.

The approach (215) has been extensively utilized by the Purdue group to model mesoscopic systems with the equivalent Landauer formula for nonequilibrium Green's functions. For mesoscopic waveguides in the linear response regime, even with dissipation present, they have shown that this form can be extended to the use of a Wigner distribution, which then can be used to define a local electrochemical potential, but that sensible results are obtained when these potentials are defined in an average sense over regions the size of a de Broglie wavelength.[398] They have also investigated the simulation of dissipative scatterers by the use of voltage probes, but find that the formulation is robust enough to introduce the scatterers directly into the Green's function.[399] This latter has led to a general multi-probe formula[400] and to a.c. response.[401] In Fig. 43, the solutions for the electrochemical potential and current density are shown for a quantum waveguide with a stub sidearm. One important result of this calculation is the large potential drops at the cathode and anode contacts apparent in the structure.

The role of surface scattering in mesoscopic quantum waveguides has also been investigated through the use of the Kubo formulation.[402] In this latter work, it is found that the presence of edge roughness can open gaps in the energy spectrum of the carriers for a one-dimensional wire, but this does not occur in higher dimensions. The Green's function approach has also been utilized to study magnetotransport in these quasi-one dimensional wires.[403,404]

Usually, the Kubo formula and the Green's functions developed above have been limited to the nearly (or exactly) homogeneous semiconductor. Jauho has extended this approach to the strongly inhomogeneous situation by the introduction of a new ansatz.[405] Here, the Green's function $G^<(\mathbf{k}_{xy}, \Delta z, \omega, \mathbf{R})$ is expanded as a product of the spectral density and a quasi-distribution, the latter of which is composed of a mini-(Wigner)distribution for each transverse mode present at $\mathbf{R}$ and the wave functions propagating in the z-direction as

$$G^<(\mathbf{k}_{xy}, \Delta z, \omega, \mathbf{R}) = \sum_{\lambda,\eta,\eta'=\pm} \psi^*_{\lambda,\eta}(R_z + \frac{\Delta z}{2})\psi_{\lambda,\eta'}(R_z - \frac{\Delta z}{2}) \times f^W_{\eta\eta'}(\mathbf{k}_{xy}, E_\lambda, \mathbf{R}) A(\mathbf{k}_{xy}, E_\lambda, \mathbf{R}) \,. \tag{217}$$

Here, the inhomogeneous potential along a mesoscopic device may be treated non-perturbatively. Others have looked at other generalizations of (215),[56] and at the time-dependent transport in nonlinear structures.[406] Finally, there has been an effort to extend the general approach to obtain e.g. an energy balance equation along the line of the treatment of Secs. 7 and 8.[407]

The formula (216) has found significant usage at low temperatures to model quantum wires with what are known as real-space lattice Green's functions, or recursive Green's functions.[408,409] These were studied initially in connection with universal conductance fluctuations,[80] and have since been used extensively to study general conductance in multi-terminal mesoscopic devices.[85,410–412] Some formal considerations on extending an arbitrary Hamiltonian result to more general Hamiltonians has also been presented.[413] It has also been extended to the consideration of Coulomb interactions among the electrons on the lattice sites,[414] and to self-consistent fields due to gated potentials.[415]

The general approach to the site-representation evaluation of the Kubo formula begins with writing a new Hamiltonian for the total "energy" in the system. This may be done as

$$H = \sum_{j,k} E_{jk} c^\dagger_{jk} c_{jk} + \sum_{j,k} \{ V_x [c^\dagger_{jk} c_{j-1,k} + c^\dagger_{j-1,k} c_{jk}] + V_y [c^\dagger_{jk} c_{j,k-1} + c^\dagger_{j,k-1} c_{jk}] \} \,. \tag{218}$$

This leads to an equivalent expression for the current in the x-direction as

$$j_x = i \frac{2\pi e a}{h} \sum_{j,k} V_x [c^\dagger_{jk} c_{j-1,k} - c^\dagger_{j-1,k} c_{jk}] \,, \tag{219}$$

and the conductivity is then, from the Kubo formula

$$\sigma_{xx} = -\frac{4V_x^2 L_x^2 e^2}{hV} Tr \left\{ \frac{1}{E - H + i\eta} \sum_k [c_{jk}^\dagger c_{j-1,k} - c_{j-1,k}^\dagger c_{jk}] \right\} .$$

$$\times \sum_{k'} [c_{jk'}^\dagger c_{j-1,k'} - c_{j-1,k'}^\dagger c_{jk'}] \frac{1}{E - H + i\eta} \right\} . \qquad (220)$$

Then, this may be simplified by recognizing that the energy functions are the Green's functions, and

$$\sigma = -\frac{4V_x^2 L_x^2 e^2}{hV} Tr \{ G''_{kk'}(j'-1,j) G''_{k'k}(j-1,j') + G''_{kk'}(j',j) G''_{k'k}(j-1,j'-1)$$

$$- G''_{kk'}(j'-1,j) G''_{k'k}(j-1,j') - G''_{kk'}(j',j-1) G''_{k'k}(j-1,j'-1) \} . \qquad (221)$$

In the above, the quantities V_x, L_x, V and E_{jk} are the interaction energy between neighboring sites, the length of the "lattice" in the x-direction, the "volume" of the simulation crystal, and the site energy, respectively. Usually, the interaction energy is held fixed, at least for one direction, and the site energy incorporates the disorder (due to impurity scattering, for example) by allowing this quantity to vary over the range $-W/2$ to $W/2$ with a uniform distribution—the so-called Anderson model. The trace runs over the subscripts k and k' in the y-direction. Here, we will use the notation that $G(j)$ is a column matrix whose elements are $G_k(j)$.

It is necessary to build up the overall Green's function, and this is usually a self-consistent process. Rather than follow this usual case, an approximation is made by propagating metallic Green's functions in from the boundary as

$$G^L(j) = \{ G_0^{-1}(j) + |V_x|^2 G^L(j-1) \} , \qquad (222)$$

$$G^R(j) = \{ G_0^{-1}(j) + |V_x|^2 G^R(j+1) \} , \qquad (223)$$

$$G_{0,k}(j) = \frac{1}{E - E_{jk} + i\eta} . \qquad (224)$$

The Green's functions are propagated separately from the left and the right directions, and then the two versions are added together to give the total Green's function. Once this is done, the conductivity can be computed as a function of the energy E.

Variations of this general approach have been used by Sols *et al.* to study a stub-tuned waveguide "transistor."[416] In Fig. 44, the general structure of the device, and the transmission coefficient for a range of parameters, are shown. The approach has also been used to study the resonant-tunneling diode in the presence of electron-phonon coupling.[119,417] Fig. 45 shows the general structure of the device and the tunneling coefficient as a function of energy in the presence of the dissipative interaction. The density in the quantum well is illustrated in Fig. 46 as a function of the transmission coefficient, while the build-up time is plotted as a function of the well width in Fig. 47.

20. The Resonant-Tunneling Diode

Section 4 above discussed the variety of devices that have become the test vehicles for examining the role of quantum transport. Foremost among these is the resonant tunneling diode. A variation of this, the quantum dot, in which a resonant section of an electron waveguide is coupled to two other (input/output) waveguides, was also discussed. It turns out that these are essentially the only devices in which any extensive study of quantum transport with the Green's functions has been made. One approach in fact uses a single particle Green's function, and the poles of this latter function, to identify the resonant states within the resonant-tunneling diode.[418] Using the Kubo formula, it is conceptually easy to determine the current through the resonant-tunneling diode, although the introduction of a self-consistent potential distribution complicates this task. A quasi-analytic theory to determining the self-consistent population of the resonant state was introduced by Lu and Horing.[419] Here, it was found that there could be a suppression of the tunneling current, due to the exclusion principle, when the resonant state was significantly occupied. On the other hand, Davies *et al.*[420] have demonstrated that the semi-classical rate equation picture holds both in the coherent limit, in which transport through the double barrier is described in a simple wave mechanics picture, and in the limit where scattering in the resonant state makes a classical "sequential tunneling" picture more appropriate. In the case where the electrons are incident with an energy greater than that of the resonant state, elastic scattering can "focus" the electrons into final states whose final states for motion perpendicular to the barrier planes closely matches the energy of the resonant state.[421] This implies that the lateral momentum contributions are randomized in such a manner that the longitudinal momentum closely matches the energy of the resonant level.

The role of inelastic scattering in the resonant tunneling diode has been examined by a variety of authors within the Green's function formalism. The earliest studies focused upon the tunneling of carriers which were assisted by an infrared photon beam, with the result that the photon-induced resonant tunneling could be switched and controlled by the change of the applied bias.[422] Later work showed that the role of the inelastic processes depended upon the transparency of the barriers, particularly in thicker barriers where the charge in the well would be larger.[423] In Fig. 48, the density in the resonant-tunneling diode is shown as a plot of the distribution function versus both energy and position. Here, it is clear that a phonon emission process is underway during the tunneling of the particles through the structure. The possible role of multiple scattering (higher-order interactions) has also been studied recently.[424]

The tunneling time for carriers through the double-barrier structure has been of some interest in connection with the debate over the tunneling time for particles moving through a barrier. One approach is to calculate the ohmic conductance through the structure as a function of the frequency of the bias potential.[425] From these studies, it is found that the basic a.c. conductance rolls off at a frequency corresponding to the width of the resonance level ΔE_r divided by Planck's constant, or

$$f_{ro} = \frac{\Delta E_r}{\hbar} \, . \tag{225}$$

However, these latter authors find that there is a peak in the conductance if the barriers are unequal, which in fact is the usual case. Other studies indicate that lateral confinement also affects the a.c. conductance,[426] as does the presence of charge in the well,[427] where the absence of charge in the well basically slows the transient switching due to the need to charge the well and its intermediate state. A more complete theory is one in which it is necessary to discuss the differences between transits through the well, charging of the well, and tunneling processes.[428] From this latter, it is demonstrated that there are two characteristic time scales of importance—that of the steady-state current characteristic time (which appears in the conductance) and that of the charge in the well Q_W. These two define the dwell time and the transit time, where the latter is defined by Q_W/j_{dc}. The transit time is an intrinsically static quantity, while the response time (similar to the roll-off frequency discussed above) is important for a.c. properties. The response time is found to be dominated by the escape rates out of the quantum well for the stored charge in the well. Carrier-carrier scattering does not affect these conclusions to any great extent.[429]

As pointed out in the Introduction, the quantum dot is a mesoscopic equivalent of the resonant tunneling diode, except that it is usually studied in the few electron regime. The mesoscopic quantum dot has been studied in analogy to that of the diode. Here, in an approach that treats the charging of the dot region through a Hubbard-like term in the Hamiltonian, it is found that charging of the dot region strongly affects the overall transmission.[430] Of course, the few-electron quantum dot is subject to the normal single-electron charging with conductance peaks as electrons charge and discharge the dot region.[431,432] The overall behavior has recently been reviewed[433] and will not be dealt with to any great extent here. The use of Green's functions has, however, identified the importance of carrier-carrier interactions,[434] and the nonlinearities introduced by the charge in the well.[435]

The use of the real-time Green's functions has mainly been limited in practice to nearly homogeneous high field structures and to linear response in inhomogeneous structures. Nevertheless, it has the capability to provide the most in depth look at the actual physics of nanodevices, and its use is still open to considerable study.

VI. Acknowledgements

The authors are indebted to a great number of people who have contributed to this effort and provided permission to use their own figures and results. Our own work, along with our students and collaborators, has been supported by ONR, ARO, and AFOSR. Unpublished algorithm development and results on the density matrix was developed along with T. R. Govindam (SRA, Inc.), who also contributed to the formulation of the understanding of the formulation of dissipation in the density matrix. Major discussions over the years with A.-P. Jauho (Technical University of Denmark), W. R. Frensley (University of Texas at Dallas), S. Goodnick (Oregon State University), U. Ravaioli and K. Hess (University of Illinois), J. R. Barker (Glasgow

University), A. M. Kriman (SUNY Buffalo), M. Büttiker and R. Landauer (IBM Research), R. Bertoncini (CRS4 Cagliari, Italy), and L. Reggiani and C. Jacoboni (University of Modena), plus a multitude of students, have contributed greatly to our level of understanding of this field, and the authors are properly cognizant of the debt owed to these people.

References

[1] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, IEEE J. Sol. State Circ. **9**, 256 (1974).

[2] G. Baccarani, M. R. Wordeman, and R. H. Dennard, IEEE Trans. Electron Dev. **31**, 452 (1984)

[3] G. G. Shahidi, D. A. Antoniadis, and H. I. Smith, J. Vac. Sci. Technol. B **6**, 137 (1988).

[4] G. A. Sai-Halasz, M. R. Wordeman, D. P. Kern, S. Rishton, and E. Ganin, IEEE Electron Dev. Lett. **9**, 464 (1988).

[5] Y. Jin, D. Mailly, F. Carcenac, B. Etiene, and H. Launois, Microelectr. Engr. **6**, 195 (1987).

[6] R. C. Tiberio, E. D. Wolf, S. F. Anderson, W. J. Schaff, P. J. Tasker, and L. F. Eastman, J. Vac. Sci. Technol. B **6**, 134 (1988).

[7] G. Bernstein, D. K. Ferry, and E. Newman, Superlatt. Microstruc. **4**, 308 (1985); IEEE Electron Dev. Letters **10**, 444 (1986); Superlatt. Microstruc. **5**, 373 (1986); IEEE Trans. Electron Dev. **35**, 887 (1988); J. Ryan, J. Han, A. Kriman, D. K. Ferry, and P. Newman, Sol.-State Electron. **32**, 1609 (1989).

[8] J. Han, D. K. Ferry, and P. Newman, IEEE Electron Dev. Lett. **11**, 209 (1990).

[9] A. Ishibashi, K. Funato, and Y. Mori, Jpn. J. Appl. Phys. 27, L2382 (1988).

[10] D. K. Ferry, in *Granular Nanoelectronics*, Ed. by D. K. Ferry, J. R. Barker, and C. Jacoboni (Plenum Press, New York, 1990) 1.

[11] D. K. Ferry, J. R. Barker, and C. Jacoboni, Eds., *Nonlinear Transport in Semiconductors* (Plenum Press, New York, 1979).

[12] J. R. Barker and D. K. Ferry, Sol.-State Electron. **23**, 519 (1980); **23**, 531 (1980); **23**, 545 (1980).

[13] M. A. Reed, Ed., *Nanostructured Systems* (Academic Press, New York, 1992).

[14] H. A. Cerdeira, F. Guinea López, and U. Weiss, Eds., *Quantum Fluctuations in Mesoscopic and Macroscopic Systems* (World Scientific Press, Singapore, 1990).

[15] B. L. Altshuler, P. A. Lee, and R. A. Webb, Eds., *Mesoscopic Phenomena in Solids* (North-Holland, Amsterdam, 1991).

[16] C. M. Snowden, Intern. J. Microwave and Millimeter-Wave Comp.-Aided Eng. **1**, 4 (1991).

[17] S. E. Laux and M. V. Fischetti, IEEE Electron Dev. Lett. **9**, 467 (1988).

[18] C. M. Snowden, Ed., *Semiconductor Device Modeling* (Springer-Verlag, London, 1989).

[19] T. Ando, F. Stern, and A. B. Fowler, Rev. Mod. Phys. **54**, 437 (1982).

[20] U. Ravaioli and D. K. Ferry, IEEE Trans. Electron Dev. **33**, 677 (1986).

[21] P. J. Price, Superlatt. Microstruc. **2**, 213 (1986).

[22] E. Kobayashi, C. Hamaguchi, T. Matsuoka, and K. Taniguchi, IEEE Trans. Electron Dev. **36**, 2353 (1989).

[23] M. B. Patil and U. Ravaioli, Sol.-State Electron. **33**, 953 (1990).

[24] R. Drury, R. E. Miles, and C. M. Snowden, in *Proc. Intern. Conf. on Computational Electronics*, in press; R. K. Veresegyhazy and C. M. Snowden, in *Proc. Intern. Conf. on Computational Electronics*, in press.

[25] H. Tanimoto, N. Yasuda, K. Taniguchi, and C. Hamaguchi, Jpn. J. Appl. Phys. **27**, 563 (1988).

[26] X. L. Lei, J. Phys. C. **18**, L593 (1985).

[27] S. E. Laux and F. Stern, Appl. Phys. Lett. **49**, 91 (1986).

[28] M. B. Patil and U. Ravaioli, Superlatt. Microstruc. 8, 459 (1990); T. Kerkhoven, A. T. Glick, U. Ravaioli, J. H. Arends, and Y. Saad, J. Appl. Phys. **68**, 3461 (1990).

[29] G. V. Chester, Rept. Prog. Phys. **26**, 411 (1963).

[30] J. R. Barker, J. Phys. C **6**, 2773 (1973).

[31] K. K. Thornber, Sol.-State Electron. **21**, 259 (1978).

[32] D. K. Ferry and J. R. Barker, J. Phys. Chem. Sol. **41**, 1083 (1980).

[33] D. Lowe, J. Phys. C **18**, L209 (1985).

[34] P. Lipavský, F. S. Khan, F. Abdolsalami, and J. W. Wilkins, Phys. Rev. B **43**, 4885 (1991).

[35] See, e.g., J. R. Barker, in *Granular Nanoelectronics*, Ed. by D. K. Ferry, J. R. Barker, and C. Jacoboni (Plenum Press, New York, 1990) 1.

[36] See, e.g., the discussion in C. George, I. Prigogine, and L. Rosenfeld, Nature **240**, 25 (1972), and W. H. Zurek, Phys. Today, October 1991, p. 36.

[37] P. Lipavský, V. Špička, and B. Velický, Acta Phys. Polon. **A73**, 319 (1988).

[38] See the discussion in R. Landauer, Physics Today **31**, 23 (November 1978).

[39] See, e.g., B. Misra, Proc. Natl. Acad. Sci. USA **75**, 1627 (1978).

[40] E. Wigner, Phys. Rev. **40**, 749 (1932).

[41] J. G. Kirkwood, J. Chem. Phys. **14**, 180 (1946).

[42] See, e.g., P. N. Argyres and S. Sfiat, Phys. Lett. A **146**, 231 (1990), and references contained therein.

[43] See, e.g., the discussion in D. E. Aspnes, Phys. Rev. **147**, 554 (1966).

[44] A review is found in M. A. Stroscio, in *Introduction to Semiconductor Technology: GaAs and Related Compounds*, Ed. by C. T. Wang (John Wiley, New York, 1990) 551.

[45] A. M. Kriman, N. C. Kluksdahl, and D. K. Ferry, Phys. Rev. B **36**, 5953 (1987).

[46] G. J. Iafrate, H. L. Grubin, and D. K. Ferry, J. Physique (Colloq., C-10) **42**, 307 (1981).

[47] D. Bohm, Phys. Rev. **85**, 166 (1952); Phys. Rev. **85**, 180 (1952).

[48] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965) Sec. 11.1-2.

[49] R. P. Feynman and H. Kleinert, Phys. Rev. A **34**, 5080 (1986); H. Kleinert, Phys. Lett. B **181**, 324 (1986).

[50] D. K. Ferry and J. R. Zhou, Phys. Rev. B **48**, 7944 (1993).

[51] See, e.g., U. Rössler, in *Quantum Coherence in Mesoscopic Systems*, Ed. by B. Kramer (Plenum Press, New York, 1991) 45.

[52] H. Sakaki, Jpn. J. Appl. Phys. **19**, L735 (1980); Inst. Phys. Conf. Ser. **63**, 251 (1982); J. Vac. Sci. Technol. **19**, 148 (1981).

[53] These are reviewed by M. Heiblum, in *High Speed Electronics*, Ed. by B. Kallback and H. Beneking (Springer-Verlag, Berlin, 1986), and J. R. Hayes, A. J. F. Levi, A. C. Gossard, and J. H. English, in *High Speed Electronics*, Ed. by B. Kallback and H. Beneking (Springer-Verlag, Berlin, 1986).

[54] R. Landauer, in *Nonlinearity in Condensed Matter*, Ed. by A. R. Bishop, D. K. Campbell, P. Kumar, and S. E. Trullinger (Springer-Verlag, Berlin, 1987) 2.

[55] R. Landauer, IBM J. Res. Develop. **1**, 223 (1957).

[56] Y. Meir and N. S. Wingreen, Phys. Rev. Lett. **68**, 2512 (1992).

[57] S. Namba, C. Hamaguchi, and T. Ando, Eds., *Science and Technology of Mesoscopic Structures* (Springer-Verlag, Tokyo, 1992).

[58] R. Landauer and J. W. F. Woo, Phys. Rev. B **5**, 1189 (1972).

[59]Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).

[60]D. Yu. Sharvin and Yu. V. Sharvin, JETP Lett. **34**, 272 (1981).

[61]B. Pannetier, J. Chaussy, R. Rammal, and P. Gandit, Phys. Rev. Lett. **53**, 718 (1984).

[62]R. A. Webb, S. Washburn, C. Umbach, and R. A. Laibowitz, Phys. Rev. Lett. **54**, 2696 (1985); S. Washburn, C.P. Umbach, R. B. Laibowitz, and R. A. Webb, Phys. Rev. B **32**, 4789 (1985).

[63]V. Chandrasekhar, M. J. Rooks, S. Wind, and D. E. Prober, Phys. Rev. Lett. **55**, 1610 (1985).

[64]M. Büttiker, Annals New York Acad. Sci. **480**, 194 (1985).

[65]C. P. Umbach, S. Washburn R. B. Laibowitz, and R. A. Webb, Phys. Rev. B **30**, 4048 (1984).

[66]J. C. Licini, D. J. Bishop, M. A. Kastner, and J. Melngailis, Phys. Rev. Lett. **55**, 2987 (1985).

[67]R. G. Wheeler, K. K. Choi, A. Goel, R. Wisnieff, and D. E. Prober, Phys. Rev. Lett. **49**, 1674 (1982).

[68]S. B. Kaplan and A. Harstein, Phys. Rev. Lett. **56**, 2403 (1986).

[69]P. A. Lee, A. D. Stone, and H. Fukuyama, Phys. Rev. B **35**, 1039 (1987).

[70]M. Büttiker, Physica B **175**, 199 (1991); Phys. Rev. Lett. **68**, 843 (1992).

[71]M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, Phys. Rev. B **31**, 6207 (1985); M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986); M. Büttiker, Phys. Rev. B **32**, 1846 (1986).

[72]W. J. Skocpol, P. M. Mankiewich, R. E. Howard, L. D. Jackel, D. M. Tennant, and A. D. Stone, Phys. Rev. Lett. **56**, 2865 (1986).

[73]W. J. Skocpol, in *Physics and Fabrication of Microstructures and Microdevices,* Ed. by M. J. Kelly (Springer-Verlag, London, 1986) 255.

[74] R. Landauer, in *Analogies in Optics and Micro Electronics*, Ed. by W. van Haeringen and D. Lenstra (Kluwer Academic Publ., Amsterdam, 1990) 243.

[75] R. Landauer, Z. Phys. B **68**, 217 (1987).

[76] L. F. Register, U. Ravaioli, and K. Hess, J. Appl. Phys. **69**, 7153 (1991).

[77] V. Chandrasekhar, D. E. Prober, and P. Santhanam, Phys. Rev. Lett. **61**, 2253 (1988); V. Chandrasekhar, P. Santhanam, and D. E. Prober, Phys. Rev. B **44**, 11203 (1991).

[78] R. B. Griffiths, Phys. Rev. Lett. **70**, 2201 (1993).

[79] M. Büttiker, Phys. Rev. B **33**, 3020 (1986).

[80] D. S. Fisher and P. A. Lee, Phys. Rev. B **23**, 685 (1981).

[81] M. Büttiker, IBM J. Res. Develop. **32**, 317 (1988); in *Electronic Properties of Multilayers and Low-Dimensional Semiconductor Structures*, Ed. by J. M. Chamberlain et al. (Plenum Press, New York, 1990) 51.

[82] R. Landauer, Physica Scripta **T42**, 110 (1992).

[83] P. J. Price, Sol. State Commun. **84**, 119 (1992); Appl. Phys. Lett. **62**, 289 (1993).

[84] M. Büttiker and H. Thomas, Inst. Phys. Conf. Series **127**, 19 (1992); M. Büttiker, A. Prêtre, and H. Thomas, Phys. Rev. Lett. **70**, 4114 (1993).

[85] H. U. Baranger, in *Computational Electronics*, Ed. by K. Hess, J. P. Leburton, and U. Ravaioli (Kluwer, Norwell, MA, 1991) 201.

[86] B. J. van Wees, H. van Houten, C. W. J. Beenakker, J. G. Williamson, L. P. Kouwenhouven, D. van der Marel, and C. T. Foxon, Phys. Rev. Lett. **60**, 848 (1988).

[87] D. A. Wharam, U. Ekenberg, M. Pepper, D. G. Hasko, H. Ahmed, J. E. F. Frost, D. A. Ritchie, D. C. Peacock, and G. A. C. Jones, Phys. Rev. B **39**, 6283 (1989).

[88] M. Büttiker, Phys. Rev. Lett. **62**, 229 (1989); in *Nanostructure Physics and Fabrication*, Ed. by M. A. Reed and W. P. Kirk (Academic Press, New York, 1989) 319; in *Nanostructured Systems*, Ed. by M. A. Reed (Academic Press, New York, 1992) 191.

[89] M. Büttiker, Phys. Rev. B **38**, 9375 (1988).

[90] P. Streda, J. Kucera, and A. H. MacDonald, Phys. Rev. Lett. **59**, 1973 (1987).

[91] T. Martin and S. Feng, Phys. Rev. Lett. **64**, 1971 (1990).

[92] A. M. Chang, Sol. State Commun. **74**, 871 (1990).

[93] F. Sols, in *Phonons in Semiconductor Nanostructures*, Ed. by J. P. Leburton, J. Pascual, and C. M. Sotomayor-Torres (Kluwer Academic Publ., 1992).

[94] L. W. Molenkamp and M. J. M. de Jong, to be published.

[95] R. Landauer, J. Phys. Cond. Matter **1**, 8099 (1989).

[96] G. Y. Hu and R. F. O'Connell, J.Phys. Cond. Matter **4**, 9623 (1992).

[97] S. Ho and K. Yamaguchi, Semicond. Sci. Technol. **7**, B430 (1992).

[98] M. Büttiker, Phys. Rev. B **40**, 3409 (1989).

[99] M. Büttiker, Ann. NY Acad. Sci. **581**, 176 (1990); Festkorperprobleme **30**, 41 (1990).

[100] M. Büttiker, Phys. Rev. B **38**, 13297 (1988).

[101] G. Müller, D. Weiss, K. von Klitzing, K. Ploog, H. Nickel, W. Schlapp, and R. Lösch, Phys. Rev. B **46**, 4336 (1992).

[102] M. Macucci, U. Ravaioli, and T. Kerkhoven, Superlatt. Microstruc. **12**, 509 (1992); U. Ravaioli, T. Kerkhoven, M. Raschke, and A. T. Galick, Superlatt. Microstruc. **11**, 343 (1992).

[103] M. A. Stroscio, G. J. Iafrate, K. W. Kim, M. A. Littlejohn, A. Bhatt, and M. Dutta, in *Integrated Optics and Optoelectronics*, Ed. by K.-K. Wong and M. Razeghi, *Crit. Reviews of Opt. Science and Technology* CR45, 341 (1993).

[104] M. Chen, W. Porod, and D. J. Kirkner, submitted for publication

[105] L. Esaki, Phys. Rev. **109**, 603 (1958).

[106] See, e.g., E. O. Kane, J. Phys. Chem. Sol. **12**, 181 (1959), and the next two references.

[107] P. N. Argyres, Phys. Rev. **126**, 1386 (1962).

[108] P. J. Price, in *Handbook of Semiconductors*, Vol.1 (Revised), Ed. by P. T. Landsberg (North-Holland, Amsterdam, 1992).

[109] A. O. Caldeira and A. J. Leggett, Phys. Rev. Lett. **46**, 211 (1981).

[110] H. Grabert and U. Weiss, Phys. Rev. Lett. **54**, 1605 (1985).

[111] See, e.g., the review by M. Jonson, in *Quantum Transport in Semiconductors*, Ed. by D. K. Ferry and C. Jacoboni (Plenum Press, New York, 1992); recent studies have continued to be published, for example, A. P. Jauho and M. Jonson, J. Phys. Cond. Matter **1**, 9027 (1989); J. A. Støvneng and A. P. Jauho, Phys. Rev. B **47**, 446 (1993).

[112] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. **49**, 57 (1982).

[113] L. Esaki and R. Tsu, IBM J. Res. Develop. **14**, 61 (1970); R. Tsu and L. Esaki, Appl. Phys. Lett. **22**, 562 (1973).

[114] L. L.Chang, L. Esaki, and R. Tsu, Appl. Phys. Lett. **24**, 593 (1974).

[115] D. K. Ferry, in *Physics of Quantum Electronic Devices,* Ed. by F. Capasso (Springer-Verlag, Berlin, 1990) 77.

[116] T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. O. Peck, Appl. Phys. Lett. **43**, 588 (1983).

[117] See, e.g., M. Büttiker, IBM J. Res. Develop. **32**, 63 (1988).

[118] A. P. Jauho, Phys. Rev. B **41**, 12327 (1990).

[119] J. A. Støvneng, E. H. Hauge, P. Lipavský, and V. Špička, Phys. Rev. B **44**, 13595 (1991).

[120] M. Sumetskii, Phys. Lett. A **153**, 149 (1991).

[121] M. Wagner, Jpn. J. Appl. Phys., in press.

[122] F. W. Sheard and G. W. Toombs, Appl. Phys. Lett. **52**, 1228 (1988).

[123] H. P. Joosten, H. J. M. F. Noteborn, K. Kaski, and D. Lenstra, J. Appl. Phys. **70**, 3141 (1991).

[124] T. Fiig and A. P. Jauho, Appl. Phys. Lett. **59**, 2245 (1991); Surf. Sci. **267**, 392 (1992).

[125] C. Jacoboni and P. J. Price, Sol. State Commun. **75**, 193 (1990); in *Resonant Tunneling in Semiconductors*, Ed. by L. L. Chang *et al.* (Plenum Press, New York, 1991) 351.

[126] J. H. Davies, P. Hyldgaard, S. Hershfield, and J. W. Wilkins, Phys. Rev. B **46**, 9620 (1992).

[127] M. A. Reed, J. N. Randall, R. J. Aggarwal, R. J. Matyi, T. M. Moore, and A. E. Westel, Phys. Rev. Lett. **60**, 535 (1988).

[128] H. Mizuta, C. Goodings, M. Wagner, and S. Ho, J. Phys. Cond. Matter **4**, 8783 (1992).

[129] M. P. Stopa, Surf. Sci. **367**, 286 (1992); M. P. Stopa and Y. Tokura, in *Science and Technology of Mesoscopic Structures*, Ed. by S. Namba, C. Hamaguchi, and T. Ando (Springer-Verlag, Tokyo, 1992) 297; M. Stopa, submitted for publication.

[130] A. S. Sachrajda, R. P. Taylor, C. Dharma-Wardana, P. Zawadski, J. A. Adams, and P. T. Coleridge, Phys. Rev. B **47**, 6811 (1993).

[131] M. Wagner, U. Merkt, and A. V. Chaplik, Phys. Rev. B **45**, 1951 (1992).

[132] U. Merkt, J. Huser, and M. Wagner, Phys. Rev. B **43**, 7320 (1991).

[133] See, e.g., Z. Shao, W. Porod, and C. S. Lent, submitted for publication; W. Porod, Z. Shao, and C. S. Lent, Phys. Rev. B **48**, 8495 (1993).

[134] R. P. Taylor, A. S. Sachrajda, P. Zawadski, P. T. Coleridge, and J. A. Adams, Phys. Rev. Lett. **69**, 1989 (1992).

[135] M. Sumetskii, J. Phys. Cond. Matter **3**, 2651 (1991).

[136] Y. Feng, A. S. Sachrajda, R. P. Taylor, J. A. Adams, M. Davies, P. Zawadski, P. T. Coleridge, D. Landheer, P. A. Marshall, and R. Barber, Appl. Phys. Lett. **63**, 1666 (1993).

[137] See, e.g., D. V. Averin and K. K. Likharev, in *Mesoscopic Phenomena in Solids*, Ed. by B. L. Altshuler, P. A. Lee, and R. A. Webb (North-Holland, Amsterdam, 1991) 173; K. K. Likharev, in *Granular Nanoelectronics*, Ed. by D. K. Ferry, J. R. Barker, and C. Jacoboni (Plenum Préss, New York, 1992).

[138] L. J. Geerligs, V. F. Anderegg, P. A. M. Holweg, J. Mooij, H. Pothier, D. Esteve, C. Urbina, and M. H. Devoret, Phys. Rev. Lett. **64**, 2691 (1990).

[139] H. Pothier, P. Lafarge, P. F. Orfila, C. Urbina, D. Esteve, and M. H. Devoret, Physica B **169**, 573 (1991); Europhys. Lett. **64**, 2691 (1992).

[140] M. P. Stopa, Surf. Sci. **263**, 433 (1992); J. P. Bird, K. Ishibashi, Y. Aoyagi, and T. Sugano, Sol.-State Electron., in press.

[141] M. Sumetskii, Phys. Rev. B **46**, 4702 (1992).

[142] J. D. White and M. Wagner, Phys. Rev. B **48**, 2799 (1993).

[143] J. W. Wilkins, S. Hershfield, J. H. Davies, P. Hyldgaard, and C. J. Stanton, Phys. Scripta **T42**, 115 (1992); S. Hershfield, J. H. Davies, P. Hyldgaard, C. J.Stanton, and J. W. Wilkins, Phys. Rev. B **47**, 1967 (1993).

[144] D. K. Ferry, Prog. Quantum Electron. **16**, 251 (1993).

[145] K. Nakazato, T. J. Thornton, J. White, and H. Ahmed, Appl. Phys. Lett. **61**, 3145 (1992); K. Nakazato, R. J. Blaikie, J. R. A. Cleaver and H. Ahmed, Electron. Lett. **29**, 384 (1993).

[146]H. Matsuoka, T. Ichiguchi, T. Yoshimura, and E. Takeda, IEDM Conf. Proc. (IEEE Press, New York, 1992) 781.

[147]K. Yano and D. K. Ferry, Superlatt. Microstruc. **11**, 61 (1992); Phys. Rev. B **46**, 3865 (1992).

[148]B. Kramer and J. Mašek, Z. Physik B **76**, 457 (1989).

[149]D. J. Thouless, Phys. Rev. Lett. **39**, 1167 (1977).

[150]H. Cruz and S. Das Sarma, J. Physique I **3**, 1515 (1993).

[151]R. Bertoncini, submitted for publication.

[152]F. W. J. Hekking, Yu. V. Nazarov, and G. Schön, Europhys. Lett. **14**, 489 (1991), **20**, 255 (1992).

[153]F. Sols, in *Nanostructures and Mesoscopic Systems*, Ed. by M. A. Reed (Academic Press, New York, 1992) 417; in *Phonons in Semiconductor Nanostructures*, Ed. by J. P. Leburton, J. Pascaul, and C. M. Sotomayor-Torres (Kluwer, Norwall, CT, 1992).

[154]L. F. Register and K. Hess, submitted for publication.

[155]N. S. Wingreen, K. W. Jacobsen, and J. W. Wilkins, Phys. Rev. Lett. **61**, 1396 (1988).

[156]W. Cai, T. F. Zheng, P. Hu, B. Yudanin, and M. Lax, Phys. Rev. Lett. **63**, 418 (1989).

[157]H. A. Fertig, S. He, and S. Das Sarma, Phys. Rev. B **41**, 3596 (1990).

[158]S. Datta, M. Cahay, and M. McLennan, Phys. Rev. B **36**, 5655 (1987); M. Cahay, M. McLennan, and S. Datta, Phys. Rev. B **37**, 10125 (1988); H. R. Frohne, M. J. McLennan, and S. Datta, J. Appl. Phys. **66**, 2699 (1989).

[159]M. Macucci and K. Hess, Phys. Rev. B **46**, 15357 (1992).

[160]Y. Takagaki and D. K. Ferry, Phys. Rev. B **45**, 8506 (1992).

[161]Y. Takagaki and D. K. Ferry, Phys. Rev. B **44**, 8399 (1991).

[162] F. Sols, M. Macucci, U. Ravaioli, and K. Hess, in *Nanostructure Physics and Fabrication*, Ed. by M. A. Reed (Academic Press, New York, 1989) 157.

[163] A. Weisshaar, J. Lary, S. M. Goodnick, and V. K. Tripathi, Appl. Phys. Lett. **55**, 2114 (1992); SPIE **1284**, 45 (1990).

[164] Y. Takagaki and D. K. Ferry, Phys. Rev. B **45**, 6715 (1992).

[165] A. Weisshaar, J. Lary, S. M. Goodnick, and V. K. Tripathi, IEEE Electron Dev. Lett. **12**, 2 (1991).

[166] Y. Takagaki and D. K. Ferry, Phys. Rev. B **45**, 13494 (1992).

[167] A. Weisshaar, J. Lary, S. M. Goodnick, and V. K. Tripathi, J. Appl. Phys. **70**, 355 (1991).

[168] M. Macucci, U. Ravaioli, and T. Kerkhoven, Superlatt. Microstruc. **12**, 509 (1992).

[169] Y. Takagaki and D. K. Ferry, Phys. Rev. B **45**, 12152 (1992); J. Appl. Phys. **72**, 5001 (1992); J. Phys. Cond. Matter **5**, 1975 (1993).

[170] Y. Takagaki and D. K. Ferry, J. Phys. Cond. Matter **4**, 10421 (1992); Phys. Rev. B **46**, 15218 (1992).

[171] F. Sols and M. Macucci, Phys. Rev. B **41**, 11887 (1990).

[172] J. C. Wu, M. N. Wybourne, W. Yindeepol, A. Weisshaar, and S. M. Goodnick, Appl. Phys. Lett. **59**, 102 (1991); J. C. Wu, M. N. Wybourne, A. Weisshaar, and S. M. Goodnick, SPIE **1676**, 40 (1992); W. Yindeepol, A. Chin, A. Weisshaar, S. M. Goodnick, J. C. Wu, and M. N. Wybourne, in *Nanostructures and Mesoscopic Systems* (Academic Press, New York, 1992) 139; J. C. Wu, M. N. Wybourne, A. Weisshaar, and S. M. Goodnick, J. Appl. Phys. **74**, 4590 (1993).

[173] G. Y. Hu and R. F. O'Connell, Phys. Rev. B **42**, 1290 (1990).

[174] I. K. Marmorkos and S. Das Sarma, Phys. Rev. B **48**, 1544 (1993); S. Das Sarma, in *Topics in Condensed Matter Physics* (Nova Science, New York, 1993) in press.

[175] B. Y.-K. Hu and S. Das Sarma, Phys. Rev. Letters **68**, 1750 (1992); Phys. Rev. B **48**, 5469 (1993).

[176] M. G. Ancona, Superlatt. Microstruc. 7, 119 (1990); in *Proc. Computational Electronics* (Univ. Illinois Press, Urbana, IL, 1992).

[177] W. Pötz and J. Zhang, Phys. Rev. B **45**, 11496 (1992).

[178] M. G. Ancona and G. J. Iafrate, Phys. Rev. B **39**, 9536 (1989).

[179] U. Fano, Rev. Mod. Phys. **29**, 74 (1957).

[180] See, e.g., F. Rossi, R. Brunetti, and C. Jacoboni, in *Hot Carriers in Semiconductor Nanostructures* (Academic Press, New York, 1992) 153.

[181] R. Zwanzig, J. Chem. Phys. **33**, 1338 (1960).

[182] H. Mori, Prog. Theor. Phys. **34**, 399 (1966).

[183] W. Kohn and J. M. Luttinger, Phys. Rev. **108**, 590 (1957).

[184] J. L. Siegel and P. N. Argyres, Phys. Rev. **178**, 1016 (1969); P. N. Argyres, in *Lectures in Theoretical Physics*, Vol VIII-A (Univ. Colorado Press, Bolder, 1966) 183.

[185] P. N. Argyres and J. L. Sigel, Phys. Rev. Lett. **31**, 1397 (1973)

[186] A. Arnold, P. A. Markowich, and N. Mauser, in *Mathematical Models and Methods in Applied Sciences*, Vol. 1 (World Sci. Press, Singapore, 1991) 83.

[187] G. A. Baker, Phys. Rev. **109**, 2198 (1958).

[188] J. E. Moyal, Proc. Cambridge Phil. Soc. **45**, 99 (1949).

[189] T. B. Smith, J. Phys. A **11**, 2179 (1978).

[190] A. Janussis, A. Streklas, and K. Vlachos, Physica **107A**, 587 91981).

[191] A. Royer, Phys. Rev. A **43**, 44 (1991).

[192] E. A. Remler, Annals Phys. **95**, 455 (1975).

[193] J. R. Barker and S. Murray, Phys. Lett. **93A**, 271 (1983).

[194] A. Arnold, P. A. Markowich, and N. Mauser, Techn. Report 142 (Purdue Univ., 1991); A. Arnold and P. A. Markowich, in *Applied and Industrial Mathematics*, Ed. by R. Spigler (Kluwer Acad. Publ., Norwall, CT, 1991) 301.

[195] A. Arnold, P. Degond, P. A. Markowich, and H. Steinrück, Appl. Math. Lett. **2**, 187 (1989); A. Arnold and P. A. Markowich, Bolletino U.M.I. **4-B**, 449 (1990).

[196] A. Arnold, private communication.

[197] P. Carruthers and F. Zachariesen, Rev. Mod. Phys. **20**, 245 (1983).

[198] D. N. Zubarev, Sov. Phys. Uspekhi **3**, 320 (1960).

[199] R. A. Craig, J. Math. Phys. **9**, 605 (1968).

[200] L. V. Keldysh, Sov. Phys.-JETP **20**, 1018 (1965).

[201] L. P. Kadanoff and G. Baym, *Quantum Statistical Mechanics* (Benjamin, New York, 1962).

[202] A. Blandin, A. Nourtier, and D. W. Hone, J. Physique **37**, 369 (1976).

[203] M. Wagner, Phys. Rev. B **44**, 6104 (1991); **45**, 11595 (1992).

[204] J. Rammer and H. Smith, Rev. Mod. Phys. **58**, 323 (1986).

[205] H. Schoeller, Annals Phys., in press.

[206] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957); Can. J. Phys. **34**, 1274 (1957).

[207] See, e.g., B. Kramer and J. Mašek, in *Quantum Fluctuations in Mesoscopic and Macroscopic Systems*, Ed. by H. a. Cerdeira, F. G. López, and U. Weiss (World Sci. Press, Singapore, 1991) 3; B. Kramer, J. Mašek, V. Špička, and B. Velický, Surf. Sci. **229**, 316 (1990).

[208] See, e.g., P. N. Argyres and D. G. Resendes, J. Phys. Cond. Matter **1**, 7001 (1989).

[209]R. Kubo, in *Solid State Physics*, Vol. 17, Ed. by F. Seitz and D. Turnbull (Academic Press, New York, 1965) 269.

[210]See, e.g., L. Reggiani and T. Kuhn, in *Granular Nanoelectronics*, Ed. by D. K. Ferry (Plenum Press, New York, 1991) 287; T. Kuhn and L. Reggiani, Il Nuovo Cim. **14D**, 509 (1992); T. Kuhn, L. Reggiani, and L. Varani, Superlatt. Microstruc. **11**, 205 (1992); Semicond. Sci. Technol. **7**, 8495 (1992).

[211]See, e.g., B. L. Altshuler and A. G. Aronov, in *Electron-Electron Interactions in Disordered Systems*, Ed. by A. L. Efros and M. Pollak (North Holland, Amsterdam, 1985) 1.

[212]B. Velický, J. Mašek, and B. Kramer, Phys. Lett. A **140**, 447 (1989).

[213]G. D. Mahan, *Many-Particle Physics* (Plenum Press, New York, 1981).

[214]W. A. Schwalm and M. K. Schwalm, Phys. Rev. B **45**, 1770 (1992).

[215]L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).

[216]R. Zwanzig, Phys. Rev. **124**, 983 (1961); R. Zwanzig, K. S. J. Nordholm, and W. C. Mitchell, Phys. Rev. A **5**, 2680 (1972).

[217]G. W. Ford, J. T. Lewis, and R. F. O'Connell, Phys. Rev. A **36**, 1466 (1987); R. F. O'Connell and G. Y. Hu, in *Granular Nanoelectronics*, Ed. by D. K. Ferry (Plenum Press, New York, 1991) 313.

[218]G. Y. Hu and R. F. O'Connell, J. Phys. Cond. Matter **2**, 5335 (1990).

[219]D. K. Ferry, in *Physics of Nonlinear Transport in Semiconductors*, Ed. by D. K. Ferry, J. R. Barker, and C. Jacoboni (Plenum Press, New York, 1980) 577; J. Zimmermann, P. Lugli, and D. K. Ferry, Sol.-State Electron. **26**, 233 (1983).

[220]J. J. Niez and D. K. Ferry, Phys. Rev. B **28**, 889 (1983).

[221]G. Y. Hu and R. F. O'Connell, Physica A **153**, 114 (1988).

[222]G. Y. Hu and R. F. O'Connell, Physica A **149**, 1 (1988); Phys. Rev. B **38**, 1721 (1988); Phys. Rev. B **39**, 12717 (1989).

[223] J. Devreese, R. Evrard, and E. Kartheuser, Phys. Rev. B **12**, 3353 (1975); E. Kartheuser, R. Evrard, and J. Devreese, Phys. Rev. B **19**, 546 (1979).

[224] G. Y. Hu and R. F. O'Connell, Phys. Rev. B **36**, 5798 (1987); J. Phys. C **21**, 4325 (1988); Phys. Rev. B **40**, 3600 (1989); J. Phys. Cond. Matter **2**, 9381 (1990); Phys. Rev. B **44**, 3140 (1991).

[225] X. L. Lei and C. S. Ting, Phys. Rev. B **32**, 1112 (1985); L. Y. Chen, C. S. Ting, and N. J. M. Horing, Sol. State Commun. **73**, 437 (1990); Phys. Rev. B **42**, 1129 (1990).

[226] F. M. Peeters and J. Devreese, Phys. Rev. B **23**, 1936 (1981).

[227] P. N. Argyres, Phys. Rev. B **39**, 2982 (1989).

[228] A.-P. Jauho, Phys. Rev. B **32**, 2248 (1985).

[229] W. Cai, P. Hu, T. F. Zheng, B. Yudanin, and M. Lax, Phys. Rev. B **40**, 7671 (1989).

[230] X. J. Lu, N. J. M. Horing, and R. Enderlein, in *Proc. 19th Intern. Conf. on Phys. Semiconductors*, Warsaw (1988); X. J. Lu and N. J. M. Horing, Phys. Rev. B **41**, 2966 (1990).

[231] G. Y. Hu and R. F. O'Connell, Physica A **163**, 804 (1990); Phys. Rev. B **41**, 5586 (1990).

[232] J. J. Niez, K.-S. Yi, and D. K. Ferry, Phys. Rev. B **28**, 1988 (1983); J. J. Niez, Superlatt. Microstruc. **2**, 219 (1986).

[233] F. Peeters and J. T. Devreese, in *Functional Integration: Theory and Applications*, Ed. by J.-P. Antoine and E. Tirapequi (Plenum Press, New York) 303.

[234] X. L. Lei, J. Cai, and L. M. Xie, Phys. Rev. B **38**, 1529 (1988).

[235] X. L. Lei and N. J. M. Horing, in *Physics of Hot Electron Transport*, Ed. by C. S. Ting (World Scientific Press, Singapore, 1992) 1; H. L. Cui, X. L. Lei, and N. J. M. Horing, Phys. Rev. B, in press.

[236] X. L. Lei, J. L. Birman, and C. S. Ting, J. Appl. Phys. **58**, 2270 (1985).

[237] J. Cai, Commun. Theor. Phys. **13**, 249 (1989).

[238] P. Warmenbol, F. M. Peeters, and J. T. Devreese, Phys. Rev. B **37**, 4694 (1988); **39**, 7821 (1989); Sol.-State Electron. **32**, 1545 (1989).

[239] G. Y. Hu and R. F. O'Connell, J. Phys. Cond. Matter **3**, 4633 (1991).

[240] X. Lei and N. J. M. Horing, Phys. Rev. B **35**, 6281 (1987); Sol.-State Electron. **31**, 531 (1988).

[241] See, e.g., G. D. Mahan, Phys. Reports **145**, 231 (1987), and references therein.

[242] P. N. Argyres, J. Phys. Chem. Sol. **19**, 66 (1961).

[243] A. P. Jauho, Phys. Rev. B **32**, 2248 (1985).

[244] See, e.g., L. Reggiani, Physica Scripta **T23**, 218 (1988).

[245] See, e.g., L. Reggiani, P. Poli, L. Rota, R. Bertoncini, and D. K. Ferry, Phys. Stat. Sol. (b) **168**, K69 (1991); P. Poli, L. Rota, L. Reggiani, R. Bertoncini, and D. K. Ferry, Phys. Stat. Sol. (b) **176**, 203 (1993).

[246] B. Y.-K. Hu and S. Das Sarma, Phys. Rev. B **44**, 8319 (1991); Semicond. Sci. Technol. **7**, B305 (1992).

[247] J. T. Devreese and R. Evrard, ICTP Int. Report 74/93; Phys. Stat. Sol. (b) **78**, 85 (1976).

[248] P. Warmenbol, F. M. Peeters, J. T. Devreese, G. E. Alberga, and R. G. van Welzenis, Phys. Rev. B **31**, 5285 (1985); P. Warmenbol, F. M. Peeters, and J. T. Devreese, Sol.-State Electron. **31**, 771 (1988); Phys. Rev. B **38**, 12679 (1988); P. Warmenbol, F. M. Peeters, X. Wu, and J. T. Devreese, Phys. Rev. B **40**, 6258 (1989).

[249] N. J. M. Horing, H. L. Cui, and X. L. Lei, Phys. Rev. B **35**, 6438 (1987).

[250] W. Cai, M. C. Marchetti, and M. Lax, Phys. Rev. B **37**, 2636 (1988); T. F. Zheng, W. Cai, P. Hu, and M. Lax, Phys. Rev. B **40**, 1271 (1989).

[251] D. B. Tran Thoai and H. Haug, Phys. Stat. Sol. (b) **173**, 159 (1992).

[252] F. Rossi, T. Kuhn, J. Schilp, and E. Schöll, *Proc. 21st Intern. Conf. Physics of Semiconductors*, Beijing (1992); T. Kuhn and F. Rossi, Phys. Rev. Letters **69**, 977 (1992).

[253] N. N. Bogoliubov, *Problemi dinam. teorii v stat. Fiz.* (Moscow, 1946).

[254] D. N. Zubarev, *Nonequilibrium Statistical Mechanics* (Consultant's Bureau, New York, 1974).

[255] V. P. Kalashnikov, Physica **48**, 93 (1970).

[256] A. M. Kriman, J.-R. Zhou, and D. K. Ferry, Phys. Letters A **138**, 8 (1989).

[257] See, e.g., P. R. Holland, *The Quantum Theory of Motion* (Cambridge Univ. Press, 1993).

[258] D. Bohm and B. J. Hiley, Found. Phys. **11**, 179 (1981); **11**, 529 (1981); **12**, 1001 (1982); **14**, 255 (1984).

[259] J. R. Barker, Semicond. Sci. Technol., in press.

[260] See, e.g., F. Wang, C. L. Gardner, and D. G. Schaeffer, SIAM J. Appl. Math. **52**, 1076 (1992).

[261] H. L. Grubin, T. R. Govindan, J. P. Kreskovsky, and M. A. Stroscio, Sol.-State Electron. **36**, 1697 (1993).

[262] M. A. Stroscio, Superlatt. Microstruc. 2, 83 (1986).

[263] See, e.g., the discussion in D. K. Ferry, *Semiconductors* (Macmillan, New York, 1991).

[264] H. Fröhlich and V. V. Paranjape, Proc. Phys. Soc. (London) **B69**, 21 (1969).

[265] K. Bløtekjaer, IEEE Trans. Electron Dev. **17**, 38 (1970).

[266] R. Bosch and H. W. Thim, IEEE Trans. Electron Dev. **21**, 16 (1974).

[267] H. L. Grubin and J. P. Kreskovsky, Sol.-State Electron. 32, 1071 (1989).

[268] J.-R. Zhou and D. K. Ferry, in *Computational Electronics*, Ed. by K. Hess, J. P. Leburton, and U. Ravaioli (Kluwer, Norwall, CT, 1991) 63.

[269] J.-R. Zhou and D. K. Ferry, IEEE Trans. Electron Dev. **39**, 473 (1992); **39**, 1793 (1992); **40**, 421 (1993); Sol.-State Electronics **36**, 1289 (1993); Semiconductor Sci. Technolog., in press.

[270] J.-R. Zhou and D. K. Ferry, IEEE J. VLSI Design, in press.

[271] D. K. Ferry, Y. Takagaki, and J.-R. Zhou, Jpn. J. Appl. Phys., in press.

[272] H. Miyata, T. Yamada, and D. K. Ferry, Appl. Phys. Letters **62**, 2661 (1993).

[273] R. P. Feynman, *Statistical Mechanics, A Set of Lectures* (Benjamin/Cummings, Reading, MA, 1972).

[274] N. N. Bogoliubov, J. Phys. Sowjet. U. **10**, 256 (1946); M. Born and H. S. Green, Proc. Roy. Soc. London **A188**, 10 (1946); J. G. Kirkwood, J. Chem. Phys. **14**, 180 (1946); J. Yvon, Act. Sci. Ind. **542, 543** (Herman, Paris, 1937).

[275] H. Ehrenreich and M. H. Cohen, Phys. Rev. **115**, 786 (1959).

[276] E. N. Adams and P. N. Argyres, Phys. Rev. **102**, 605 (1956); **104**, 900 (1956).

[277] P. N. Argyres, J. Phys. Chem. Sol. **4**, 19 (1958); Phys. Rev. **109**, 1115 (1958); J. Phys. Chem. Sol. **8**, 124 (1959); Phys. Rev. **117**, 315 (1960).

[278] N. J. Horing and P. N. Argyres, in *Proc. Intern. Conf. Phys. Semicond., Exeter* (Inst. Phys., London, 1962) 58; P. N. Argyres and L. M. Roth, J. Phys. Chem. Sol. **12**, 89 (1959).

[279] J. B. Krieger and G. J. Iafrate, Phys. Rev. B **33**, 5494 (1986), **35**, 9644 (1987), **40**, 6144 (1989); G. J. Iafrate, J. B. Krieger, and Y. Li, in *Electronic Properties of Multilayers and Low-Dimensional Semiconductor Systems*, Ed. by J. M. Chamberlain et al. (Plenum, New York, 1990) 211.

[280] N. Sawaki and T. Nishinaga, J. Phys. C **10**, 5003 (1977).

[281] J. R. Barker, Sol.-State Electron. **21**, 267 (1978).

[282] X. L. Lei and C. S. Ting, Phys. Rev. B **30**, 4809 (1984); X. L. Lei, D. Y. Xing, M. Liu, C. S. Ting, and J. L. Birman, Phys. Rev. B **36**, 9134 (1987).

[283] F. Rossi, R. Brunetti, and C. Jacoboni, in *Granular Nanoelectronics*, Ed. by D. K. Ferry *et al.* (Plenum Press, New York, 1991); C. Jacoboni, Semicond. Sci. Technol. **B7**, 6 (1992).

[284] P. N. Argyres, Phys. Lett. A **171**, 373 (1992).

[285] H. Schoeller, Ann. Phys., to be published.

[286] F. Rossi and C. Jacoboni, Europhys. Lett. **18**, 169 (1992).

[287] R. Brunetti, C. Jacoboni, and F. Rossi, Phys. Rev. B **39**, 10781 (1989); F. Rossi and C. Jacoboni, Sol.-State Electron. **32**, 1411 (1989).

[288] H. Mizuta and C. J. Goodings, J. Phys. Cond. Matter **3**, 3739 (1991).

[289] F. Rossi and C. Jacoboni, Semicond. Sci. Technol. **B7**, 383 (1992).

[290] J. Bardeen, Phys. Rev. **49**, 653 (1936).

[291] H. L. Grubin, T. R. Govindan, B. J. Morrison, D. K. Ferry, and M. A. Stroscio, Semicond. Sci. Technol. **B7**, 360 (1992).

[292] H. L. Grubin, J. P. Kreskovsky, T. R. Govindam, and D. K. Ferry, Semicond. Sci. Technol., in press.

[293] W. R. Frensley, J. Vac. Sci. Technol. B **3**, 1261 (1985).

[294] A. Groshev, Phys. Rev. B **42**, 5895 (1990).

[295] P. N. Argyres, Phys. Rev. **132**, 1527 (1963).

[296] W. Cai, M. C. Marchetti, and M. Lax, Phys. Rev. B **34**, 8573 (1986), **35**, 1369 (1987).

[297] S. Das Sarma, J. K. Jain, and R. Jalabert, Phys. Rev. B **41**, 3561 (1990).

[298] A. O. Caldeira and A. J. Leggett, Physica **121A**, 587 (1983).

[299] H. Dekker, Phys. Rev. A **16**, 2126 (1977).

[300] H. L. Grubin and M. A. Stroscio, to be published.

[301] H. Krömer, W. Chien, J. S. Harris, Jr., and D. D. Edwall, Appl. Phys. Lett. **36**, 295 (1980).

[302] D. K. Schröder, *Semiconductor Material and Device Characterization* (John Wiley, New York, 1990), see eq. (2.5b).

[303] H. L. Grubin, T. R. Govindam, and M. A. Stroscio, Semicond. Sci. Tech., to be published.

[304] L. Mandel and E. Wolf, Rev. Mod. Phys. **37**, 231 (1965).

[305] H. Haken, H. Risken, and W. Weidlich, Z. Phys. **206**, 355 (1967).

[306] M. Lax and W. H. Louisell, IEEE J. Quantum Electron. **3**, 47 (1967).

[307] M. J. Bastiaans, Opt. Commun. **25**, 26 (1978).

[308] K. H. Brenner and J. Ojeda-Castaneda, Opti. Acta **31**, 213 (1984).

[309] M. J. Bastiaans, Appl. Opt. **19**, 192 (1980).

[310] M. Conner and Y. Li, Appl. Opt. **24**, 3825 (1985).

[311] V. Kumar and C. W. Carroll, Opt. Engr. **23**, 732 (1984).

[312] U. Ravaioli, M. A. Osman, W. Põtz, N. C. Kluksdahl, and D. K. Ferry, Physica B+C **134B**, 36 (1985); N. C. Kluksdahl, W. Põtz, U. Ravaioli, and D. K. Ferry, Superlatt. Microstruc. **3**, 41 (1987).

[313] R. P. Feynman, in *Quantum Implications; Essays in Honor of David Bohm*, Ed. by B. J. Hiley and F. David Peat (Routledge and Kegan Paul, London, 1987).

[314] I. B. Levinson, Sov. Phys.—JETP **30**, 362 (1970).

[315] H. Steinruck, Z. Ang. Math. u. Phys. **42**, 471 (1991).

[316] A. Szafer, A. M. Kriman, A. D. Stone, and D. K. Ferry, unpublished.

[317] G. F. Carrier and C. E. Pearson, *Partial Differential Equations: Theory and Techniques* (Academic, New York, 1976).

[318] Y. I. Shokin, *The Method of Differential Approximation* (Springer-Verlag, Berlin, 1983).

[319] W. R. Frensley, Phys. Rev. Lett. **57**, 2853 (1986); Phys. Rev. B **36**, 1570 (1986); in *Proc. IEDM*, 1986 (IEEE Press, New York, 1986) 571.

[320] N. C. Kluksdahl, A. M. Kriman, D. K. Ferry, and C. Ringhofer, Phys. Rev. B **39**, 7720 (1989).

[321] C. Ringhofer, *Mathematical Models and Methods of Applied Science* **2**, 91 (1992).

[322] A. Arnold, in *Proc. NUMSIM '91, Berlin*, Ed. by Gojewski, Deuflehard, and Markowich; A. Arnold and C. Ringhofer, in Computational Electronics, Ed. by U. Ravaioli and K. Hess (Univ. Illinois Press, Urbana, 1992); SIAM J. Num. Math., in press.

[323] A. Arnold and F. Nier, Math. Comp. **58**, 645 (1992).

[324] K. L. Jensen and F. A. Buot, J. Appl. Phys. **73**, 4409 (1993).

[325] N. C. Kluksdahl, C. Ringhofer, and D. K. Ferry, Superlatt. Microstruc. **4**, 127 (1988).

[326] F. A. Buot and K. L. Jensen, Phys. Rev. A **33**, 2544 (1986).

[327] D. R. Miller and D. P. Neikirk, Appl. Phys. Lett. **58**, 2803 (1991).

[328] N. C. Kluksdahl, A. M. Kriman, C. Ringhofer, and D. K. Ferry, Sol-State Electron. **31**, 743 (1988).

[329] W. R. Frensley, Rev. Mod. Phys. **62**, 745 (1990).

[330] B. Engquist and A. Majda, Math. Comput. **31**, 629 (1977).

[331] C. Ringhofer, D. K. Ferry, and N. C. Kluksdahl, Trans. Theory Stat. Phys. **18**, 331 (1989).

[332] A. Arnold, submitted for publication.

[333] H. Tsuchiya, M. Ogawa, and T. Miyoshi, IEEE Trans. Electron Dev. **38**, 1246 (1991).

[334] W. R. Frensley, Sol.-State Electron. **31**, 739 (1988).

[335] M. P. Shaw, H. L. Grubin, and P. Solomon, *The Gunn-Hilsum Effect* (Academic, New York, 1979).

[336] N. C. Kluksdahl, A. M. Kriman, D. K. Ferry, and C. Ringhofer, IEEE Electron Dev. Lett. **9**, 457 (1988).

[337] W. R. Frensley, Appl. Phys. Lett. **51**, 448 (1987); Superlatt. Microstruc. **4**, 497 (1988).

[338] N. C. Kluksdahl, Ph. D. dissertation, Arizona State University, August 1988, unpublished.

[339] H. Tsuchiya, M. Ogawa, and T. Miyoshi, Jpn. J. Appl. Phys. **31**, 745 (1992).

[340] J. R. Barker and D. K. Ferry, Phys. Rev. Lett. **42**, 1779 (1979).

[341] R. K. Mains and G. I. Haddad, J. Appl. Phys. **64**, 5041 (1988).

[342] W. R. Frensley, Sol.-State Electron. **32**, 1235 (1989); in *Computational Electronics*, Ed. by U. Ravaioli (1991).

[343] H. Haug and C. Ell, Phys. Rev. B **46**, 2126 (1992).

[344] H. C. Tso and N. J. M. Horing, Phys. Rev. B **44**, 11358 (1991).

[345] R. A. Hopfel.....

[346] H. Tsuchiya, M. Ogawa, and T. Miyoshi, IEEE Trans. Electron Dev. **39**, 2465 (1992); Jpn. J. Appl. Phys. **30**, 3853 (1991).

[347] Y. Isawa and T. Hatano, J. Phys. Soc. Jpn. **60**, 3108 (1991).

[348] T. Miyoshi, H. Tsuchiya, and M. Ogawa, IEEE J. Quantum Electron. **28**, 25 (1992).

[349]H. Tsuchiya, M. Ogawa, and T. Miyoshi, to be published.

[350]A. I. Larkin and Yu. N. Ovchinnikov, Sov. Phys. JETP 41, 960 (1975).

[351]X. L. Lei and C. S. Ting, Phys. Rev. B **32**, 1112 (1985).

[352]S. Datta, Phys. Rev. B **40**, 5830 (1989); J. Phys.:Cond. Matter **2**, 8023 (1990); Y. Lee, M. J. McLennan, and S. Datta, Phys. Rev. B **43**, 14333 (1991).

[353]G. D. Mahan and W. Hänsch, J. Phys. F **13**, L47 (1983); W. Hänsch and G. D. Mahan, Phys. Rev. B **28**, 1902 (1983); G. D. Mahan, Phys. Repts. **110**, 321 (1984), and references therein.

[354]L. Reggiani, P. Lugli, and A. P. Jauho, Phys. Rev. B **36**, 6602 (1987).

[355]G. D. Mahan, J. Phys. F **13**, L257 (1983); **14**, 941 (1984).

[356]N. J. Morgenstern-Horing, H. L. Cui, and G. Fiorenza, Phys. Rev. A **34**, 601 (1984).

[357]A. P. Jauho, Phys. Rev. B **29**, 1919 (1984).

[358]L. Reggiani, Physica **134B**, 123 (1985).

[359]P. Lipavský, V. Ŝpička, and B. Velický, Phys. Rev. B **34**, 3020 (1986).

[360]H. C. Tso and N. J. Morgenstern Horing, Phys. Rev. B **44**, 1451 (1991).

[361]A. P. Jauho, J. Phys. F **13**, L203 (1983).

[362]N. J. Morgenstern Horing and H.-L. Cui, Phys. Rev. B **38**, 10907 (1988).

[363]P. Lugli, L. Reggiani, and C. Jacoboni, Superlatt. Microstruc. **2**, 143 (1986); Physica Scripta **38**, 117 (1988).

[364]A. P. Jauho and L. Reggiani, Sol.-State Electron. **31**, 535 (1988); L. Reggiani, P. Lugli, and A. P. Jauho, J. Appl. Phys. **64**, 3072 (1988); A. P. Jauho, Sol.-State Electron. **32**, 1265 (1989).

[365]F. S. Khan, J. H. Davies, and J. W. Wilkins, Phys. Rev. B **36**, 2578 (1987).

[366] D. K. Ferry, A. M. Kriman, H. Hida, and S. Yamaguchi, Phys. Rev. Lett. 67, 633 (1991).

[367] D. K. Ferry, in *The Physics and Technology of SiO_2*, Ed. by R. A. B. Devine (Plenum, New York, 1988) 365.

[368] W. Quade, F. Rossi, and C. Jacoboni, Semiconduc. Sci. Technol. **7**, B502 (1992); W. Quade, F. Rossi, C. Jacoboni, and E.Schöll, Microelectron. Engr. **19**, 287 (1992).

[369] P. Lipavský, F. S. Khan, A. Kalvová, and J. W. Wilkins, Phys. Rev. B **43**, 6650 (1991).

[370] P. Lipavský, F. S. Khan, and J. W. Wilkins, Phys. Rev. B **43**, 6665 (1991).

[371] R. Bertoncini, A. M. Kriman, and D. K. Ferry, Phys. Rev. B **40**, 3371 (1989); **41**, 1390 (1990).

[372] R. Bertoncini and A. P. Jauho, Phys. Rev. B **44**, 3655 (1991); Semiconduc. Sci. Technol. **7**, B33 (1992).

[373] R. Bertoncini, Int. J. Mod. Phys. B **6**, 3441 (1992).

[374] R. Bertoncini and A. P. Jauho, Phys. Rev. Lett. **68**, 2826 (1992).

[375] R. L. Fork, C. H. Brito-Cruz, P. C. Becker, and C. V. Shank, Optics Lett. **42**, 112 (1979).

[376] A. Mysyroviecz, D. Hulin, A. Antonetti, A. Migus, W. T. Masselink and H. Morkoc, Phys. Rev. Lett. **56**, 2758 (1986); D. Hulin, A. Antonetti, M. Joffre, A.Migus, A. Mysyrovicz, N. Peyghambarian and H. M. Gibbs, Rev. Phys. Appl. **22**, 1269 (1987).

[377] A. von Lehmen, J. E. Zucker, J. P. Heritage, and D. S. Chemla, Opt. Lett. **11**, 609 (1986).

[378] N. Peyghambarian, S. W. Koch, M. Lindberg, B. Fluegel, and M. Joffre, Phys. Rev. Lett. **62**, 1185 (1988).

[379] W. H. Knox, D. S. Chemla, D. A. B. Miller, J. B. Stark, and S. Schmitt-Rink, Phys. Rev. Lett. **62**, 1189 (1988).

[380]C. Hirliman, J. F. Morhange, M. A. Kaneisha, A. Chevy, and C. H. Brito-Cruz, Appl. Phys. Lett. **55**, 2307 (1989).

[381]V. M. Galitsky and V. F. Yelesin, *Resonant Interaction of Electromagnetic Fields with Semiconductors* (Energoatomizdat, Moscow, 1986).

[382]R. Zimmermann, Phys. Stat. Sol. (b) **159**, 317 (1990); J. Lum. **53**, 187 (1992).

[383]H. Haug, Phys. Stat. Sol. (b) **173**, 139 (1992); L. Bányai, D. B. Tran Thoai, C. Remling, and H. Haug, Phys. Stat. Sol. (b) **173**, 149 (1992).

[384]H. C. Tso and N. J. Morgenstern Horing, Phys. Rev. B **44**, 8886 (1990).

[385]D. B. Tran Thoai and H. Haug, Phys. Rev. B **47**, 3574 (1993); Z. Physik B **91**, 199 (1993).

[386]S. Schmitt-Rink and D. S. Chemla, Phys. Rev. Lett. **57**, 2752 (1986).

[387]M. Lindberg and S. W. Koch, Phys. Rev. B **38**, 3342 (1988).

[388]T. Kuhn and F. Rossi, Phys. Rev. B **46**, 7496 (1992).

[389]P. Danielewicz, Ann. Phys. **152**, 239 (1984).

[390]P. A. Henning, Nucl. Phys. B **342**, 345 (1990).

[391]A. G. Hall, J. Phys. A **8**, 214 (1975).

[392]S. G. Tikhodeev, Sov. Phys. Dokl. **27**, 492 (1982), **27**, 624 (1982); Yu. A. Kukharenko and S. G. Tikhodeev, Sov. Phys. JETP **56**, 831 (1982);

[393]A. V. Kuznetsov, Phys. Rev. B **44**, 8721 (1991), **44**, 13381 (1991).

[394]T. Kuhn, E. Binder, F. Rossi, A. Lohner, K. Rick, P. Leisching, A. Leitenstorfer, T. Elsaesser, and W. Stolz, to be published; J. Schilp, T. Kuhn, and G. Mahler, Semiconduc. Sci. Technol. B, in press.

[395]A. A. Abrikosov, L. Gorkov, and I. Ye. Dzyaloshinskii, *Quantum Field Theoretical Methods in Statistical Physics*, 2nd. Ed. (Pergamon, Oxford, 1965).

[396] D. K. Ferry, A. M. Kriman, M. J. Kann, and R. P. Joshi, in *Monte Carlo Simulations of Semiconductors and Semiconductor Devices*, Ed. by K. Hess (Kluwer Acad. Publ., Norwell, MA, 1991).

[397] D. K. Ferry, A. M. Kriman, H. Hida, and S.Yamaguchi, Phys. Rev. Lett. **67**, 633 (1991).

[398] M. J. McLennan, Y. Lee, and S. Datta, Phys. Rev. B **43**, 13846 (1991); in *Computational Electronics*, Ed. by U. Ravaioli (Kluwer, Norwall, CT, 1991); Y. Lee, M. J. McLennan, G. Klimeck, R. K. Lake, and S. Datta, Superlatt. Microstruc. **11**, 137 (1992).

[399] S. Datta and R. K. Lake, Phys. Rev. B 44, 6538 (1991).

[400] S. Datta, Phys. Rev. B **45**, 1347 (1992); Phys. Rev. B **46**, 9493 (1992).

[401] S. Datta and M. P. Anantram, Phys. Rev. B **45**, 13761 (1992).

[402] M. K. Schwalm and W. A. Schwalm, Phys. Rev. B **45**, 6868 (1992).

[403] N. Mori, H. Momose, and C. Hamaguchi, Phys. Rev. B **45**, 4536 (1992).

[404] M. Wagner, Phys. Rev. B **45**, 11606 (1992).

[405] A. P. Jauho and O. Ziep, Physica Scripta **T25**, 329 (1989).

[406] N. S. Wingreen, A. P. Jauho, Y. Meir, Phys. Rev. B **48**, 8487 (1993).

[407] R. Lake and S. Datta, Phys. Rev. B **46**, 4757 (1992).

[408] G. Czycholl and B. Kramer, Sol. State Commun. **32**, 945 (1979).

[409] D. J. Thouless and S. Kirkpatrick, J. Phys. C **14**, 235 (1981).

[410] B. Kramer and J. Mašek, J. Phys. C **21**, L1147 (1988); J. Mašek, P. Lipavsky, and B. Kramer, J. Phys. Cond. Matter **1**, 6395 (1989); J. Mašek and B. Kramer, Z. Phys. B **75**, 37 91989); B. Kramer and J. Mašek, in *Science and Engineering of One- and Zero-Dimensional Semiconductors*, Ed. by S. P. Beaumont and C. M. Sotomajor-Torres (Plenum, New York, 1990) 155.

[411] H. Baranger, A. D. Stone, and D. P. DiVicenzo, Phys. Rev. B 37, 6521 (1988).

[412] S. He and S. Das Sarma, Sol.-State Electron. 32, 1695 (1989); Phys. Rev. B 40, 3379 (1989); Phys. Rev. B 48, 4629 (1993); S. Das Sarma and S. He, Int. J. Mod. Phys. B 7, 3375 (1993).

[413] W. A. Schwalm and M. K. Schwalm, Phys. Rev. B 37, 9524 (1988).

[414] W. Häusler, B. Kramer, and J. Mašek, Z. Phys. B 85, 435 (1991).

[415] Q. Li and D. K. Ferry, Superlatt. Microstruc. 11, 245 (1992).

[416] F. Sols, M. Macucci, U. Ravaioli, and K. Hess, Appl. Phys. Lett. 54, 350 (1989); J. Appl. Phys. 66, 3892 (1989); M. Macucci, V. Pevzner, L. F. Register, and K. Hess, in *New Concepts for Low-Dimensional Electronic Systems*, Ed. by G. Bauer, H. Heinrich, and F. Kuchar (Springer-Verlag, Heidelberg, 1993).

[417] J. A. Støvneng and E. H. Hauge, Phys. Rev. B 44, 13582 (1991).

[418] W. R. Frensley, Superlatt. Microstruc. 11, 347 91992).

[419] X. J. Lu and N. J. M. Horing, Phys. Stat. Sol. (b) 174, 27 (1992).

[420] J. H. Davies, S. Hershfield, P. Hyldgaard, and J. W. Wilkins, Phys. Rev. B 47, 4603 (1993).

[421] H. A. Fertig and S. Das Sarma, Phys. Rev. B 40, 7410 (1989).

[422] W. Cai, T. F. Zheng, P. Hu, M. Lax, K. Shum, and R. R. Alfano, Phys. Rev. Lett. 65, 104 (1990); W. Cai, T. F. Zheng, P. Hu, and M. Lax, Mod. Phys. Lett. B 5, 173 (1991).

[423] G. Neofotistos, R. Lake, and S. Datta, Phys. Rev. B 43, 2442 (1991); R. Lake and S. Datta, Phys. Rev. B 45, 6670 91992); R. Lake, G. Klimeck, and S. Datta, Phys. Rev. B 47, 6427 (1993); R. Lake, G.Klimeck, M. P. Anantram, and S. Datta, Phys. Rev. B 48, 15132 (1993).

[424] P. Hyldgaard, S. Hershfield, J. H. Davies, and J. W. Wilkins, to be published.

[425] C. Jacoboni and P. J. Price, *Frontiers in Condensed Matter Theory*, Ann. N.Y. Acad. Sci. **581**, 253 (1990); in *Resonant Tunneling in Semiconductors*, Ed. by L. L. Chang (Plenum, New York, 1991) 351.

[426] W. Cai, P. Hu, and M. Lax, Phys. Rev. B **44**, 3336 (1991).

[427] W. Cai and M. Lax, Phys. Rev. B **47**, 4096 (1993).

[428] E. Runge and H. Ehrenreich, Phys. Rev. B **45**, 9145 (1992); Ann. Phys. **219**, 55 (1992).

[429] T. Brandes, D. Weinmann, and B. Kramer, Europhys. Lett. **22**, 51 (1993).

[430] A. Groshev, T. Ivanov, and V. Valtchinov, Phys. Rev. Lett. **66**, 1082 (1991).

[431] J. Mašek and B. Kramer, PTB-Mitteilungen **101**, 327 (1991).

[432] Y. Meir, N. S. Wingreen, and P. A. Lee, Phys. Rev. Lett. **66**, 3048 (1991).

[433] G.-L. Ingold and Yu. V. Nazarov, in *Single Charge Tunneling*, Ed. by H. Grabert and M. H. Deverot (Plenum, New York, 1991) 21.

[434] W. Pfaff, D. Weinmann, W. Häusler, B. Kramer, and U. Weiss, Phys. Rev. Lett., in press; W. Häusler and B. Kramer, Phys. Rev. B, in press; T. Brandes, W. Häusler, K. Jauregui, B. Kramer, and D. Weinmann, Physica B **189**, 16 (1993).

[435] H. Schöller, Phys. Rev. Lett., in press; C. Bruder and H. Schöller, Phys. Rev. Lett., in press.

Captions

January 23, 1994

Figure 1 The Wigner distribution showing the density variation adjacent to an infinitely high potential in the space $x < 0$. The oscillations at higher momentum value arise from interference of the waves peaking closer to the interface.

Figure 2 A schematic depiction of the embedding of a device in its surroundings: boundaries, contacts, etc.

Figure 3 The potential distribution for a double-barrier resonant-tunneling diode. In (a), the potential is shown in equilibrium. The full states (shaded) on either sides of the barriers are coupled only by tunneling through the entire structure. Under bias, however (b), the resonant level is aligned with the states in the emitter, allowing them to tunnel through with a probability approaching unity.

Figure 4 The gate electrodes deplete the quasi-two-dimensional electron gas under them to create a quantum dot region in the inclosed area. The size of this area can be adjusted by depleting additional states with the "tuning" electrode. This adjusts the resonances and transmission through the structure.

Figure 5 A quantum waveguide equivalent of the DBRTD and quantum dot. (a) The wide region of the waveguide provides mode propagation at a lower energy than the thinner waveguides. This provides the equivalent of the resonant energy in the DBRTD and the dot levels in the quantum dot. (b) Computation of the current transmission through the structure leads to a negative differential conductance just as in the DBRTD. In this calculation, $W_3 = W_5 = 25nm, W_4 = 45nm$, and $W_1 = W_2 = 500nm$. The length of the center section is $25nm$ and that of the two barrier sections is 20 nm. In addition, a depletion in from the edge is assumed under bias, with $\delta_1 = 0, \delta_2 = 0.5$, $\delta_3 = 1.0$, and δ_4 =1.5 nm. The Fermi energy is 5 mV, and the temperature is assumed to be 0 K. [After Goodnick *et al.*,IEEE Electron Dev. Letters 12, 2 (1991), by permission.]

Figure 6 The density matrix, in equilibrium, for a DBRTD. Charge in the quantum well is indicated by the small peak centered in the gap.

Figure 7 The Wigner distribution function, in equilibrium, for a DBRTD. Again, charge in the quantum well is indicated by the spread in the small peak centered in the gap.

Figure 8 The path of integration for real-time Green's functions which are out of equilibrium. The tail extending downward connects to the thermal equilibrium Green's functions, where appropriate.

Figure 9 The quantum potential for the hyperbolic tangent potential cited in the text. The solution area of the overall device is 200 nm.

Figure 10 The quantum potential for a MESFET near pinchoff. The gate length is 24 nm, and the active layer is 39 nm thick, and doped to 10^{18} cm^{-3}.

Figure 11 The carrier density across the channel, under the gate, in a direction normal to the gate for a 24 nm GaAlAs/GaAs HEMT. The dashed curve is the case in which the quantum potential is set to zero.

Figure 12 The density matrix for a homogeneous system, in which the density as a function of the off-diagonal distance, or correlation distance, for GaAs with an electron density of $10^{17} cm^{-3}$.

Figure 13 (a) The charge density and potential are shown for a abrupt (non-self-consistent) potential step. The parameters are for GaAs at 0 K and a doping level of 10^{18} cm^{-3} in the region of low potential, and an order of magnitude lower in the region of high potential. (b) The density matrix for this situation.

Figure 14 (a) The charge density and self-consistent potential for the same situation as in Fig. 13. (b) The density matrix.

Figure 15 (a) The density profiles for a wide and a narrow barrier. The doping variations are the same as in Figs. 13 and 14, which provide a barrier height of 42.7 meV. (b) The density matrix for the case of the wide barrier.

Figure 16 The quantum potential (solid curve) and barrier potential (dashed curve) for the narrow barrier of Fig. 15.

Figure 17 (a) The charge density, self-consistent potential energy, and quantum potential for a heterojunction between GaAlAs and GaAs. The composition of the former material is taken to give a $0.3eV$ conduction band offset: (b) The density matrix for the charge in this heterojunction. The GaAlAs is the region $x > 100nm$.

Figure 18 A double-barrier structure with $5nm$ barriers, $0.3eV$ high, and a $5nm$ well is solved by the denstiy matrix technique. (a) The doping profile and the self-consistently determined density in the structure. (b) The self-consistently

determined potential and the quantum potential for this structure. The peak of the quantum potential in the well defines the quasi-bound state energy, as discussed in the text.

Figure 19 The double-barrier structure of Fig. 18 is immersed in a wider ($40nm$ wide) quantum well with depth of $0.15eV$. In (a), the doping and self-consistently determined density are shown, while in (b), the self-consistent potential and the quantum potential are plotted.

Figure 20 The real (a) and imaginary (b) parts of the density matrix for a uniform GaAs structure in a constant electric field of 500 V/cm.

Figure 21 A central 5 nm region is assumed to have different scattering properties from the bulk. The two cases consider i) a scattering time of 0.1 ps in the bulk and 0.01 ps in the central region, and ii) increase of the scattering time to 1 ps in the bulk (with 0.01 ps in the central region). In (a), the potential and quasi-Fermi levels are plotted, while in (b), the density is plotted for these two conditions.

Figure 22 A central region of 50 nm is assumed to have a *lower* scattering rate (1 ps versus 0.1 ps in the bulk). In (a), the potential and quasi-Fermi level are plotted, while (b) illustrates the density variation.

Figure 23 The potential (a) and density variation (b) across a single tunneling barrier, embedded in a central lightly-doped region (the parameters are discussed in the text). The parameter for the various curves is the applied bias.

Figure 24 The variation in the quasi-Fermi level for the structure and biases of Fig. 23.

Figure 25 The distribution of charge and the potential variation in equilibrium for a single-barrier structure (details are discussed in the text).

Figure 26 A comparison of the $\bar{n}$ inferred from capacitance voltage and the actual $n(x)$ determined from a self-consistent solution for the density matrix. The inset shows a plot of $\bar{x}^2/2L_D^2\beta$ vs. applied voltage.

Figure 27 It may be seen from this construction that the Wigner function has support (and is non-vanishing) in regions in which the wave functions vanish.

Figure 28 Gaussian wave packet that is interacting with resonant tunneling potential barriers. The barriers are indictated by the dark bands. (a) The incident wave packet, moving from the left toward the right, is just beginning to interact with the barriers. (b) The Gaussian wave packet during reflection. The incident and reflected components are just forming, as is the correlation around $k = 0$. Part of the packet is tunneling through the barriers. (c) The Gaussian packet after tunneling is nearly completed and most of the packet is reflected. The

tunneling packet isalso well formed.

Figure 29 The $I - V$ curve, solved self-consistently, for the DBRTD. The potential is increased to a maximum value of 0.5 V, and then decreased. A hysteresis is evident in the curves and arises from charging and discharging of the quantum well.

Figure 30 The potential distribution for the DBRTD for an applied bias of 0.22 V, which is near the peak of the $I - V$ curve in Fig. 33. Much of the potential is dropped across a cathode adjacent depletion layer, which leads to a contact resistance and quantization in the resulting triangular potential well.

Figure 31 The potential distribution for the DBRTD both with and without the spacer layers adjacent to the barriers, in the absence of any applied bias. The spacer layers are lightly-doped regions to create resistance matching between the active device and the contacts.

Figure 32 The potential distribution for the DBRTD both with and without the spacer layers adjacent to the barriers, with an applied bias of 0.4 V, a value near the valley of the $I - V$ curve. Clearly, the spacer layers have provided better matching to the boundaries, so that the depletion layer adjacent to the cathode is eliminated.

Figure 33 The $I - V$ curve, solved self-consistently, for the DBRTD with spacer layers adjacent to the barriers. The potential is increased to a maximum value of 0.5 V, and then decreased. A hysteresis is still evident in the curves and arises from charging and discharging of the quantum well. The peak-to-valley ratio is also enhanced by the presence of the spacer layers.

Figure 34 The current transient that results from switching the DBRTD from the peak current to the valley current, as indicated in the inset. The large oscillations are thought to be a coupled response due to the charge plasma oscillations and the RC behavior of the barrier region.

Figure 35 The Fourier transform of the large-signal switching behavior, for three different values of the mobility in the bulk regions. (a) The magnitude of the conductance in the negative-differential conductance regime. (b) The real part of the conductance, which is negative for frequencies below 1.5 THz. (c) The imaginary part of the conductance, which is inductive below 2 THz.

Figure 36 The potential (solid curves) and density (dashed curves) for a self-consistent Wigner function solution to the DBRTD. (a) No dissipation is included. The depletion in the cathode-adjacent layer is needed to match the cathode injection conditions, and this result would also be obtained from directly solving Schrödinger's equation. (b) The solutions when the energy-dependent LO and

acoustic phonon interactions are included. The depletion has almost been wiped out by the enhanced dissipation in the structure. (c) The solutions when contact resistance is added to the inelastic processes, with a drifted Fermi-Dirac distribution assumed in the boundaries. [After W. R. Frensley, Sol.-State Electron. **32**, 1235 (1989).]

Figure 37 The Wigner distribution for a gated quantum wire. (a) A voltage of -20 mV has been applied to the gate, which depletes the wire and creates a (tunneling) barrier in the wire. (b) A voltage of 20 mV, which enhances the conductance and creates a filled potential well in the wire, is applied to the wire. [After H. Tsuchiya *et al.*, Jpn. J. Appl. Phys. **30**, 3853 (1991).]

Figure 38 The real (a) and imaginary (b) parts of the retarded self-energy for the Airy function model in high electric fields. The reduced units are $\Omega = (\hbar\omega - eFs)/\Theta^{1/3}$, $\Theta = (eF/L)^{2/3}$.

Figure 39 The spectral function, $A(\mathrm{k}, s, \omega) = -2Im\{G_r(\mathrm{k}, s, \omega)\}$. The units and scales here are the same as in Fig. 42.

Figure 40 The distribution function (or more properly, the local density of particles) as a function of the electron energy. Parameters appropriate to Si are used. [After R. Bertoncini and A. P. Jauho, Phys. Rev. Lett. **68**, 2826 (1992).]

Figure 41 Energy distributions of electrons at different times during and after the pulse (a) for the full Bloch equation model and (b) for the semi-classical Boltzmann equation model. [After T. Kuhn *et al.*, Proc. NATO ARW on *Coherent Optical Interactions in Semiconductors*, Cambridge, 1993.]

Figure 42 The kinetic energies (a) and polarizations (b) as a function of time for a 100 fs laser pulse. [After T. Kuhn *et al.*, Proc. NATO ARW on *Coherent Optical Interactions in Semiconductors*, Cambridge, 1993.]

Figure 43 The solutions for the chemical potential in a quantum wire with a stub waveguide attached. (a) The chemical potential, and (b) the current density as a function of postion in the guide. Notice in particular the extensive fraction of the total potential drop which occurs at the contact regions. [After M. J. McLennan, Y. Lee, R. Lake, G. Neofotistos, and S. Datta, in *Computational Electronics*, Ed. by U. Ravaioli (Kluwer, Norwall, MA, 1991) 247.]

Figure 44 (a) Schematic view of the three-terminal structure used in the calculations of the waveguide transmission. (b) The squared modulus of the transmission coefficient for the first mode, as a function of the effective length of the stub $L*$ and of the electron energy E ($L_1 = L_3 = 10$ nm). [After F. Sols, M. Macucci, U. Ravaioli, and K. Hess, Appl. Phys. Lett. **54**, 350 (1989).]

Figure 45 The transmission as a function of energy (a) and the barrier structure

(b) for a double-barrier AlGaAs/GaAs structure. The transmission corresponds to 5 nm barriers (40% Al concentration) and a 6 nm well. [After J. A. Støvneng, E. H. Hauge, P. Lipavsky, and V. Špička, Phys. Rev. B 44, 13595 (1991).]

Figure 46 The time evolution of the particle density in the quantum well of a double barrier. Here, a hopping model is used to simulate the transport. The long- and short-dashed curves are the result of averaging over an energy window around the resonant energy level of varying width, where the range is defined by the reduce wave vector $\Gamma_k = 1.2 \times 10^{-3}(2\pi/a)$, and a is the lattice constant. [After J. A. Støvneng and E. H. Hauge, Phys. Rev. B 44, 13582 (1991).]

Figure 47 The build-up time for charge in the quantum well as a function of the initial wave packet width Δx. The amplitude of the barrier is taken to be four times the hopping energy between sites. [After J. A. Støvneng and E. H. Hauge, Phys. Rev. B 44, 13582 (1991).]

Figure 48 The Green's function for a double-barrier tunneling diode with inelastic scattering present. The plot illustrates the "current" density as a function of both position and energy. There is a clear transition of energy within the barrier indicating an inelastic tunneling process. [After R. Lake and S. Datta, in *NASCODE VII*, Ed. by J. J. H. Miller (Front-Range Publ., Boulder, CO, 1991).]

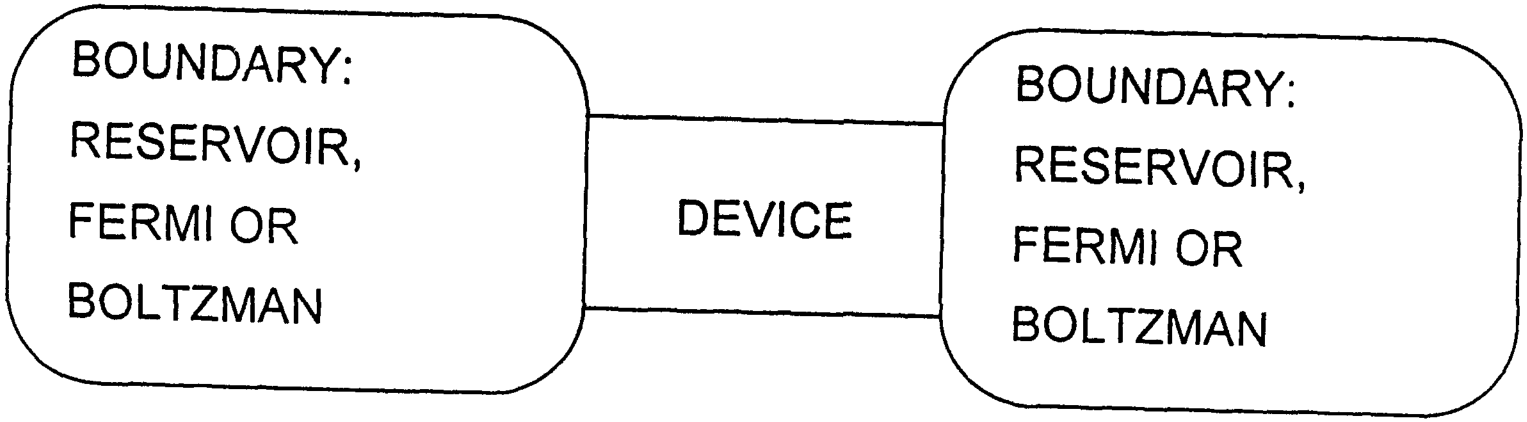

BOUNDARY:
RESERVOIR,
FERMI OR
BOLTZMAN
DEVICE
BOUNDARY:
RESERVOIR,
FERMI OR
BOLTZMAN

Fig. 2 Ferry & Grubin

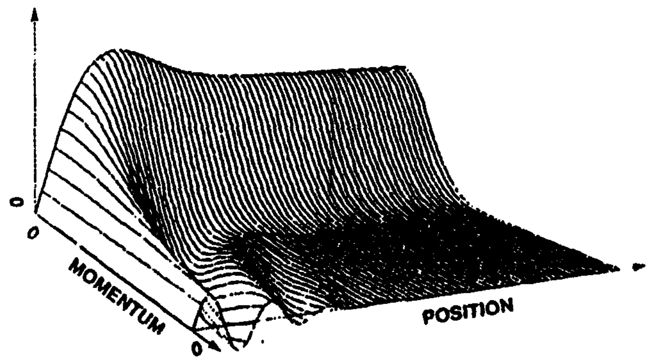
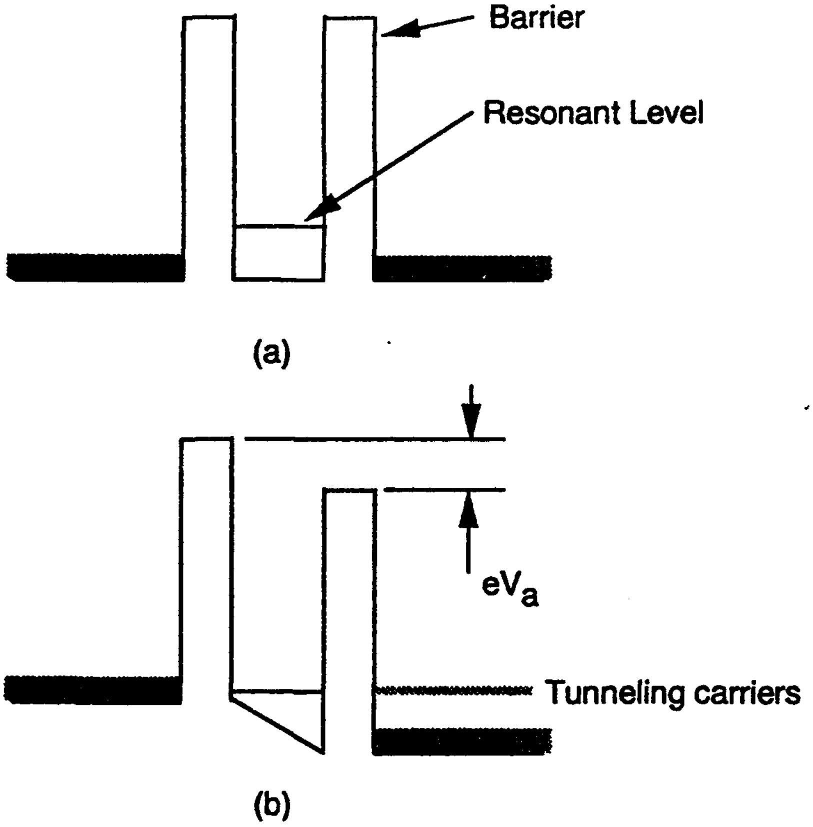


FIG. 1 Ferry & Grubin


Barrier
Resonant Level
(a)
eV_a
Tunneling carriers
(b)

[illegible]

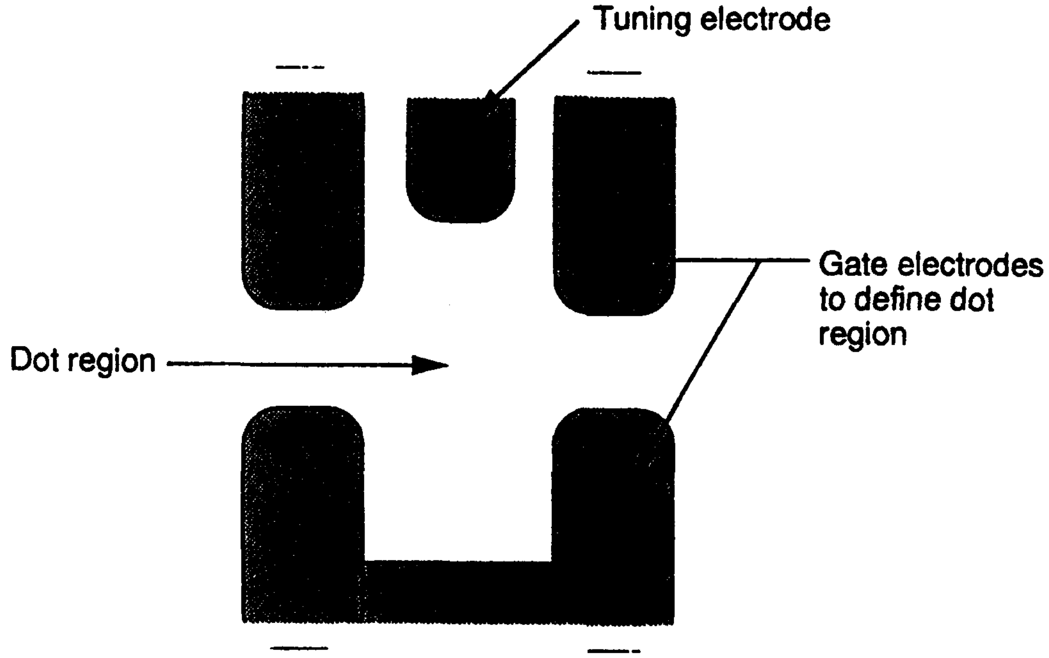
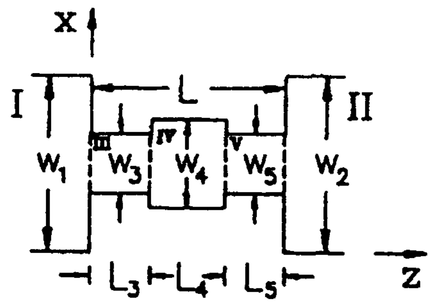
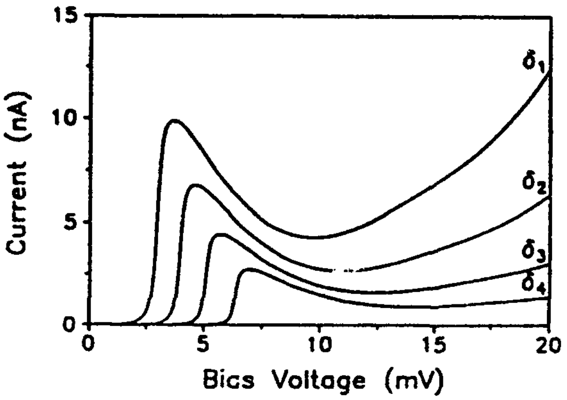

Tuning electrode
Gate electrodes to define dot region
Dot region

FIG. 4 Ferry & Grubin


X
I
II
L
III
IV
V
W_1
W_3
W_4
W_5
W_2
L_3
L_4
L_5
Z

(a)


15
10
5
0
Current (nA)
0
5
10
15
20
Bias Voltage (mV)
δ_1
δ_2
δ_3
δ_4

(b)

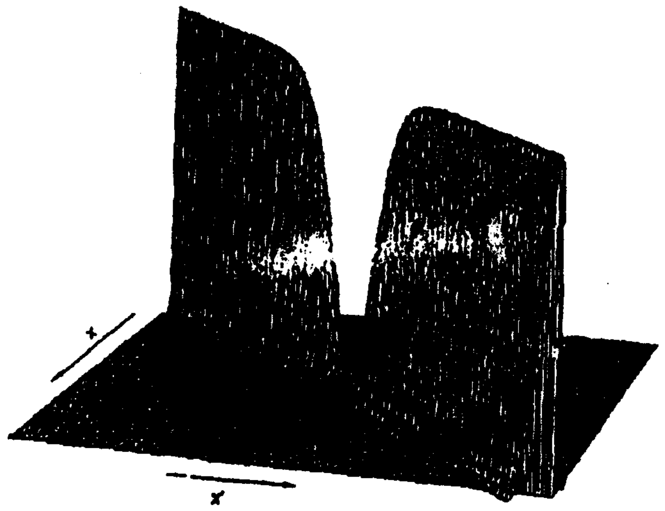


FIG. 6 Ferry & Grubin

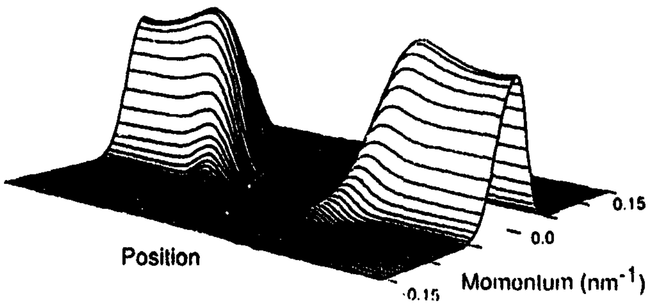


FIG. 7 Ferry & Grubin

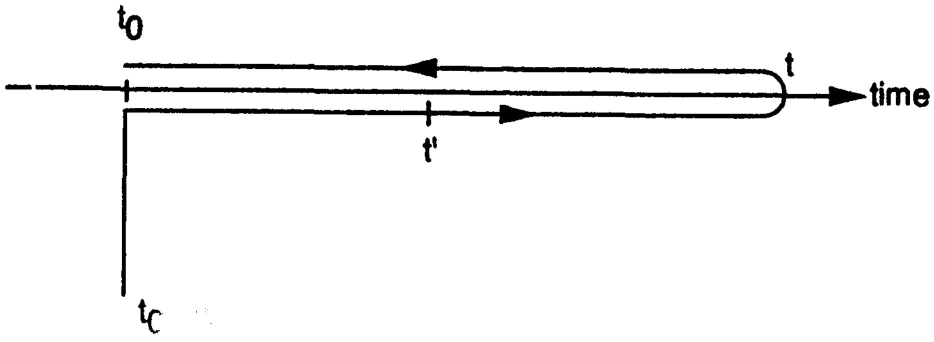

t_0
t
time
t'
t_c

FIG.8 Ferry & Grubin

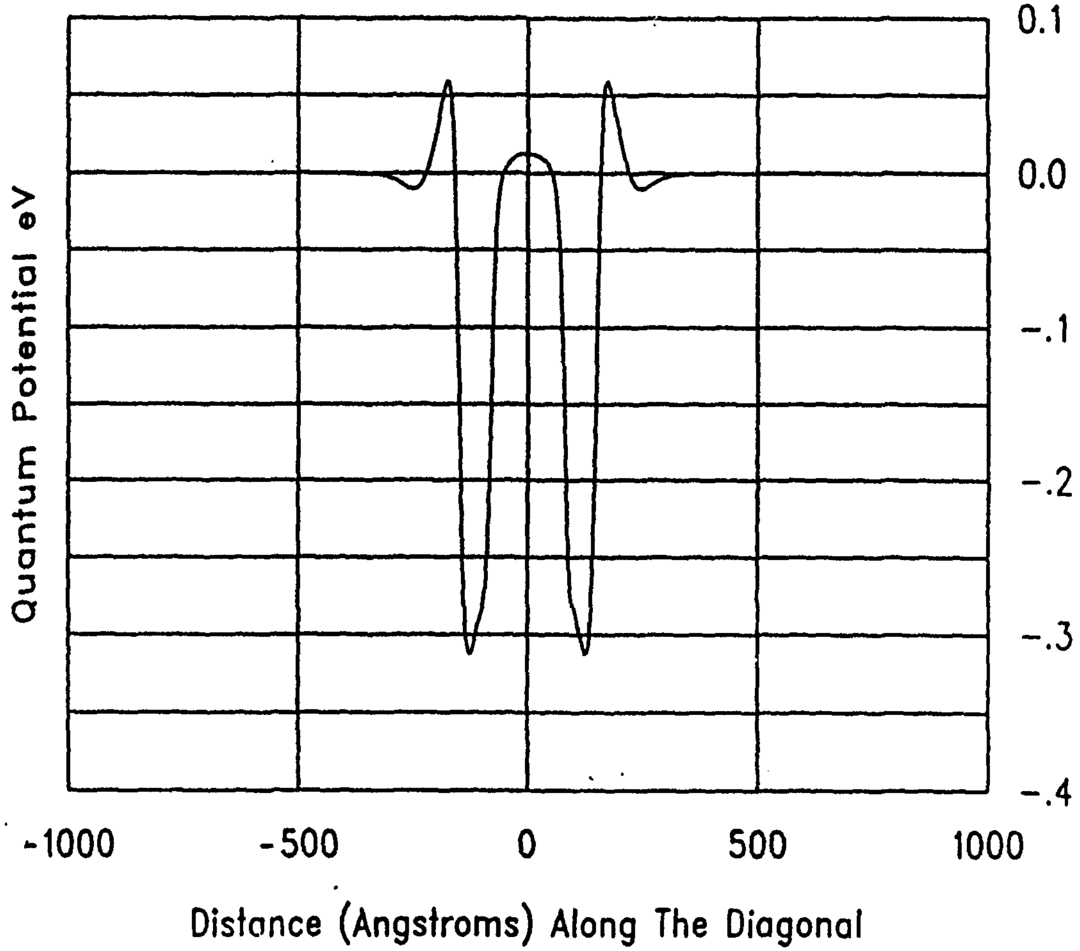


FIG. 9 Ferry & Grubin

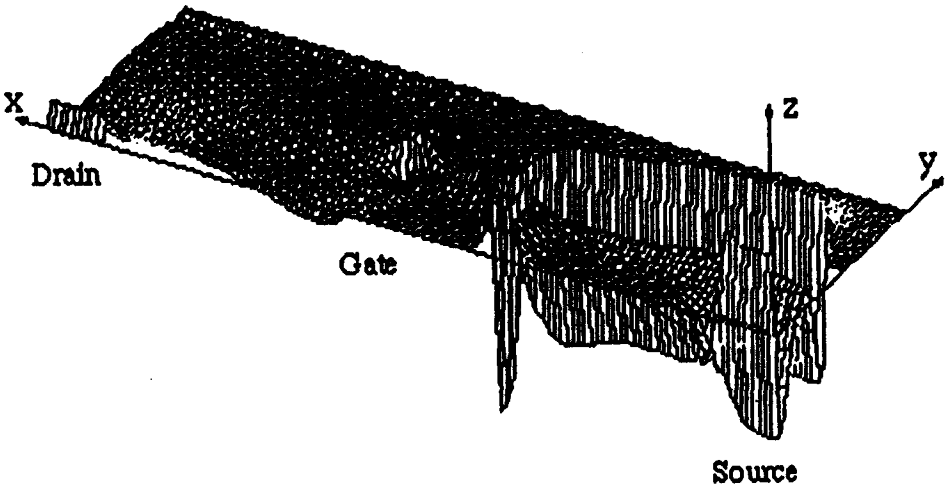


FIG.10 FERRY & GRUBIN

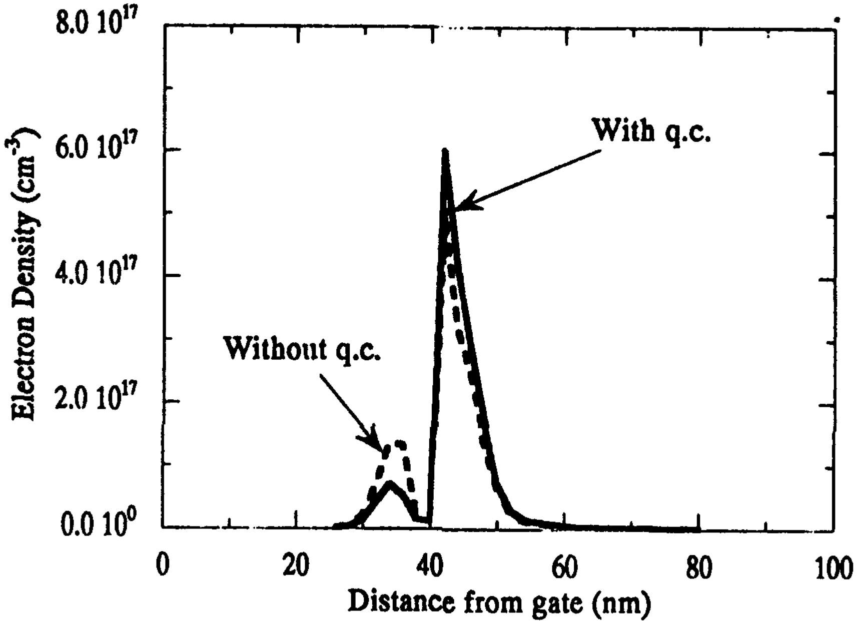


FIG. 1b Ferry & Grubin

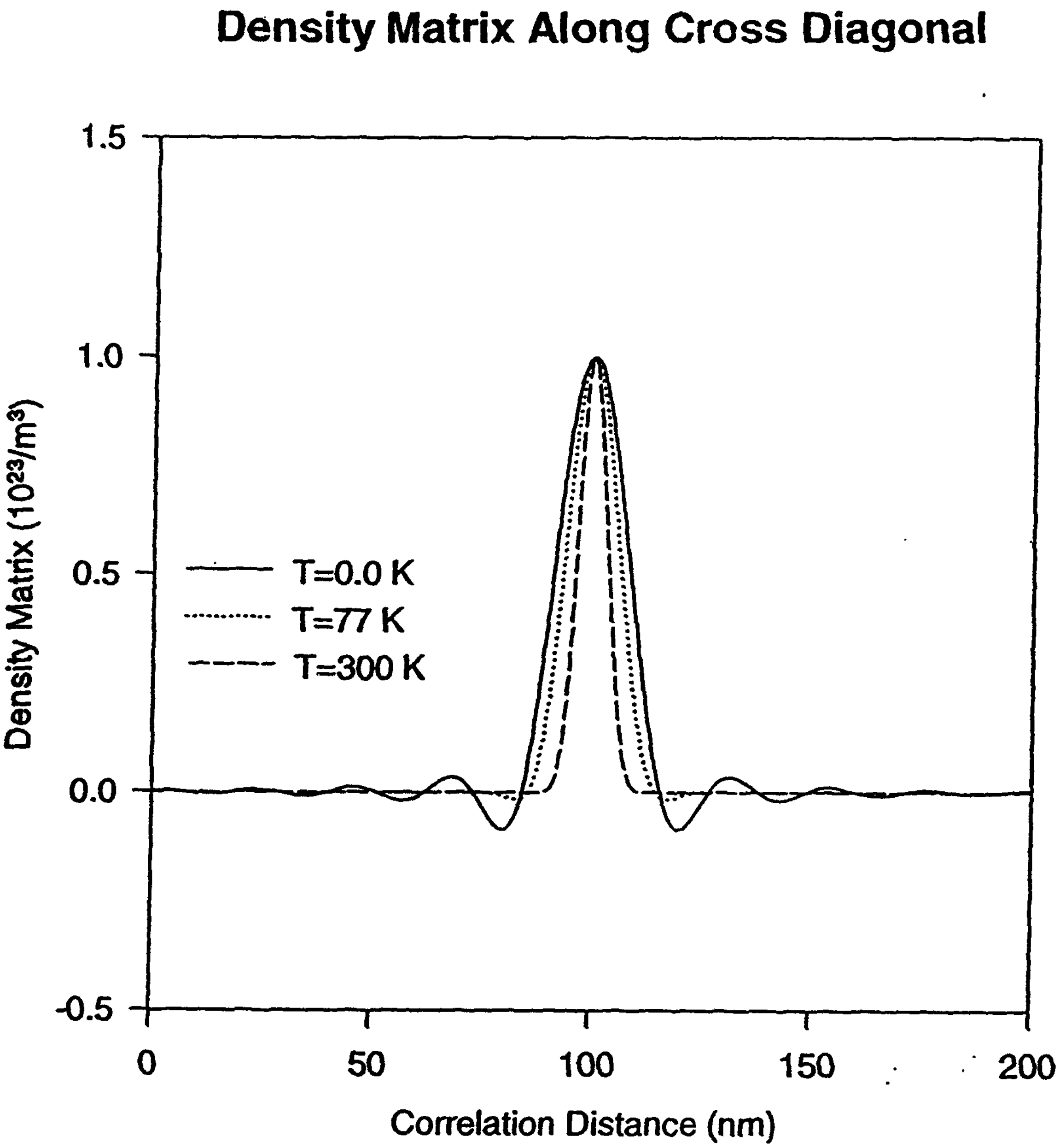


Fig. 12 Ferry & Grubin

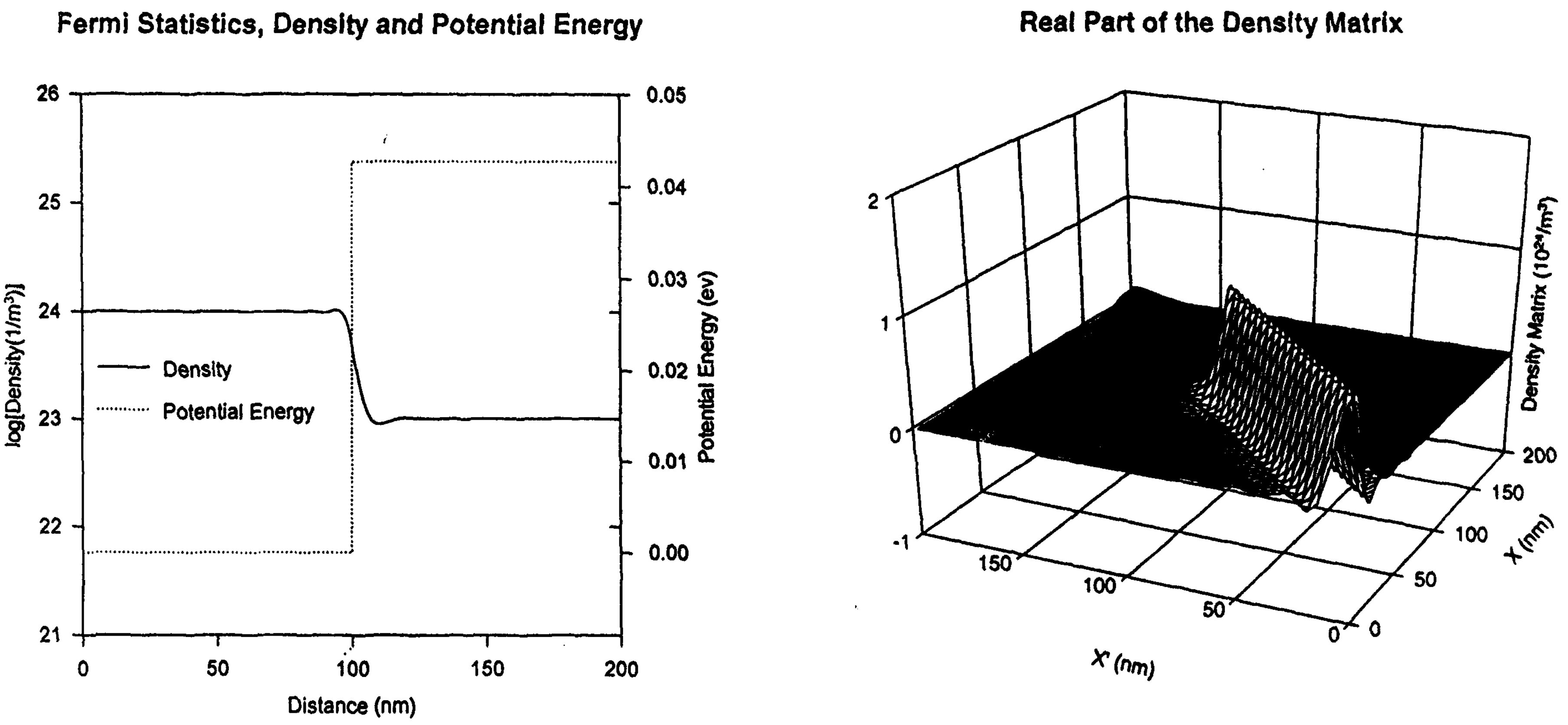
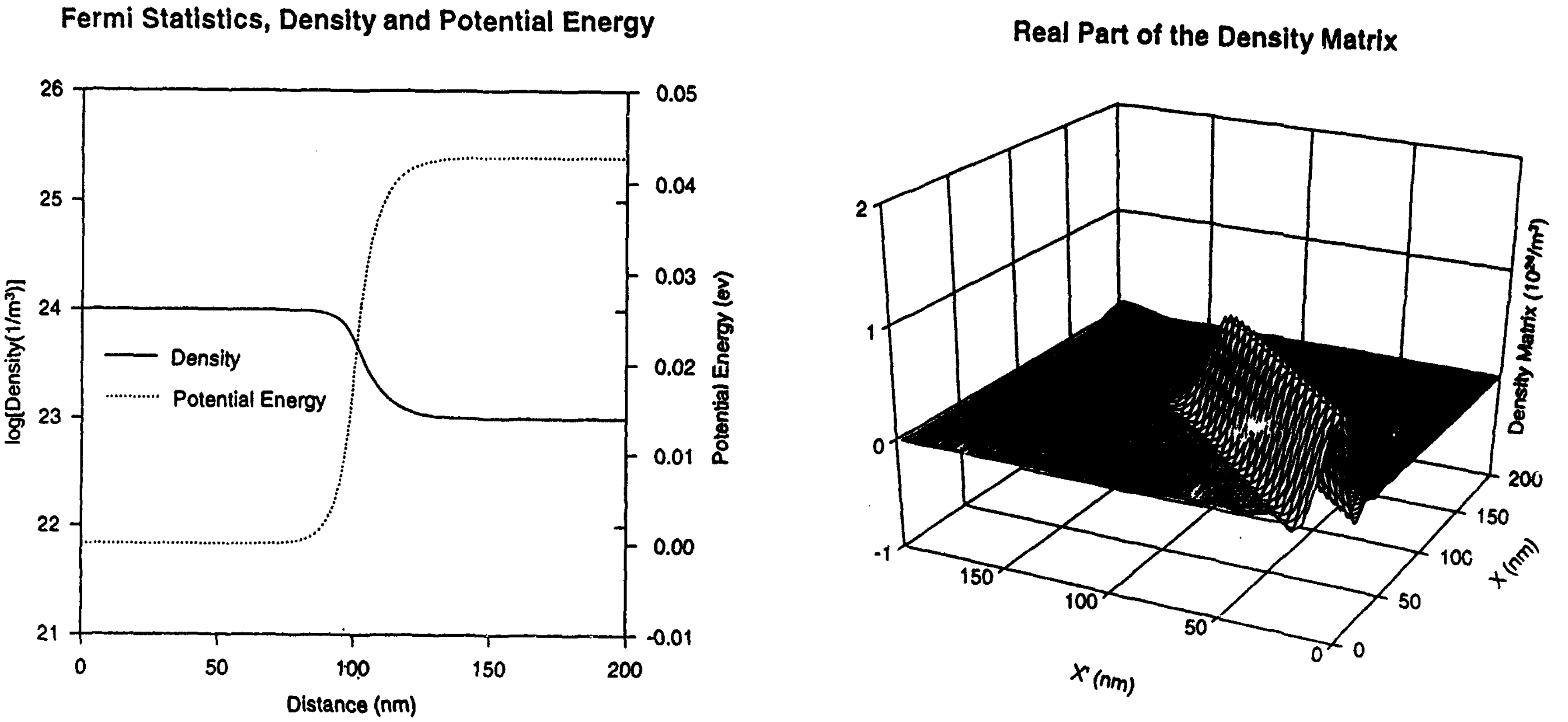


Fig. 13
Ferry & Grubin

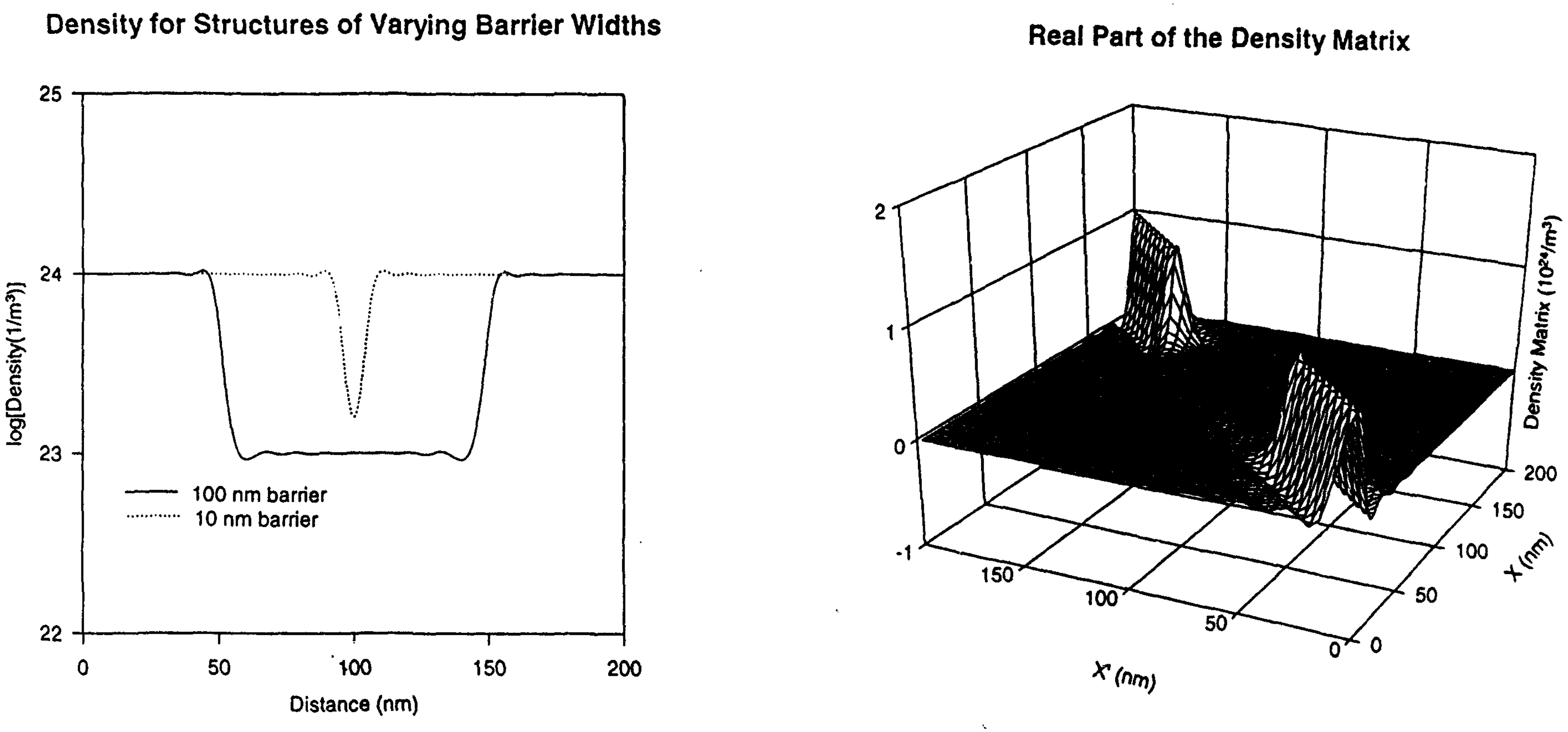


Fermi Statistics, Density and Potential Energy

Real Part of the Density Matrix

Jaques h.

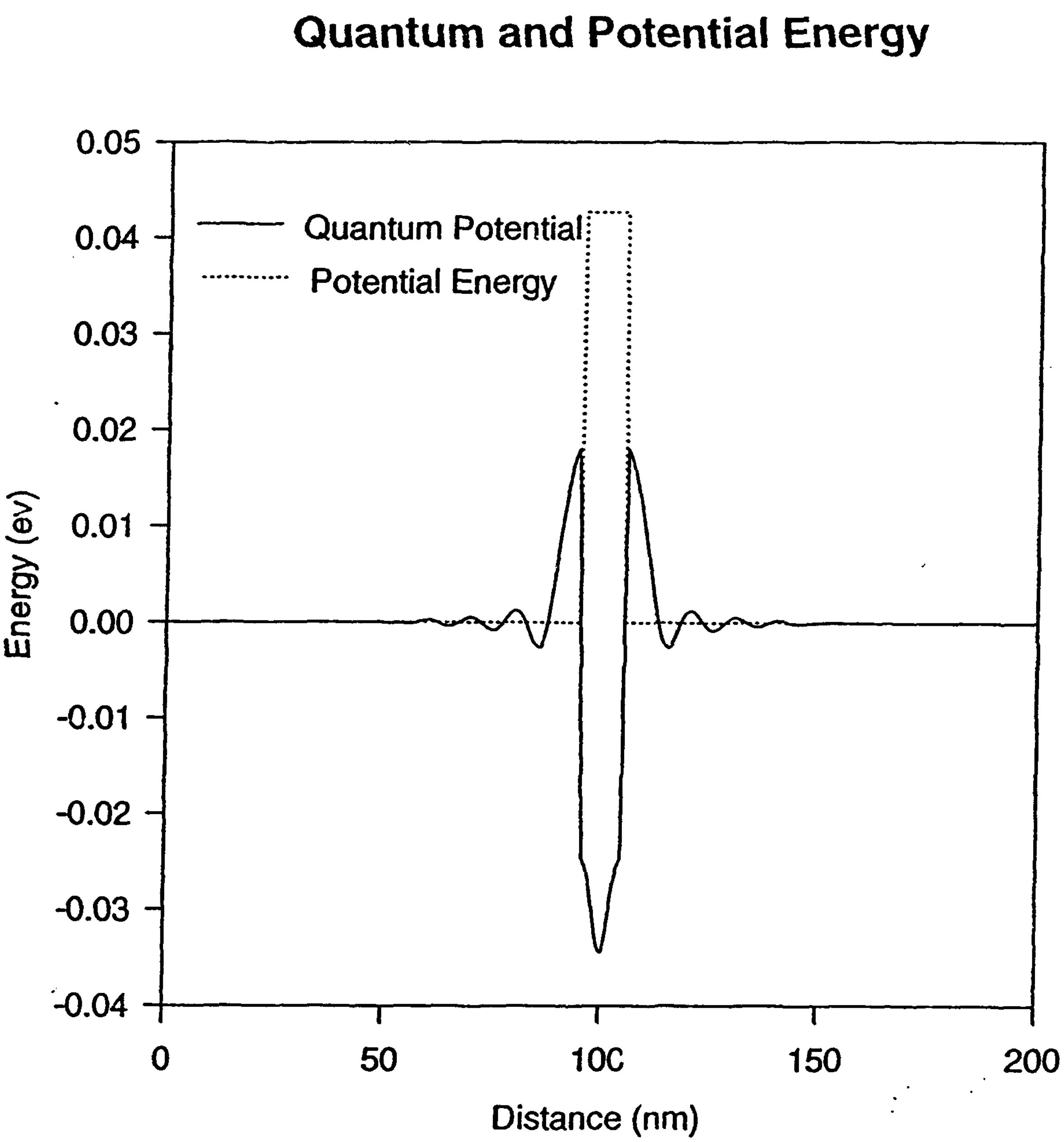
Fig. 14 Ferry & Grubin



Density for Structures of Varying Barrier Widths

Real Part of the Density Matrix

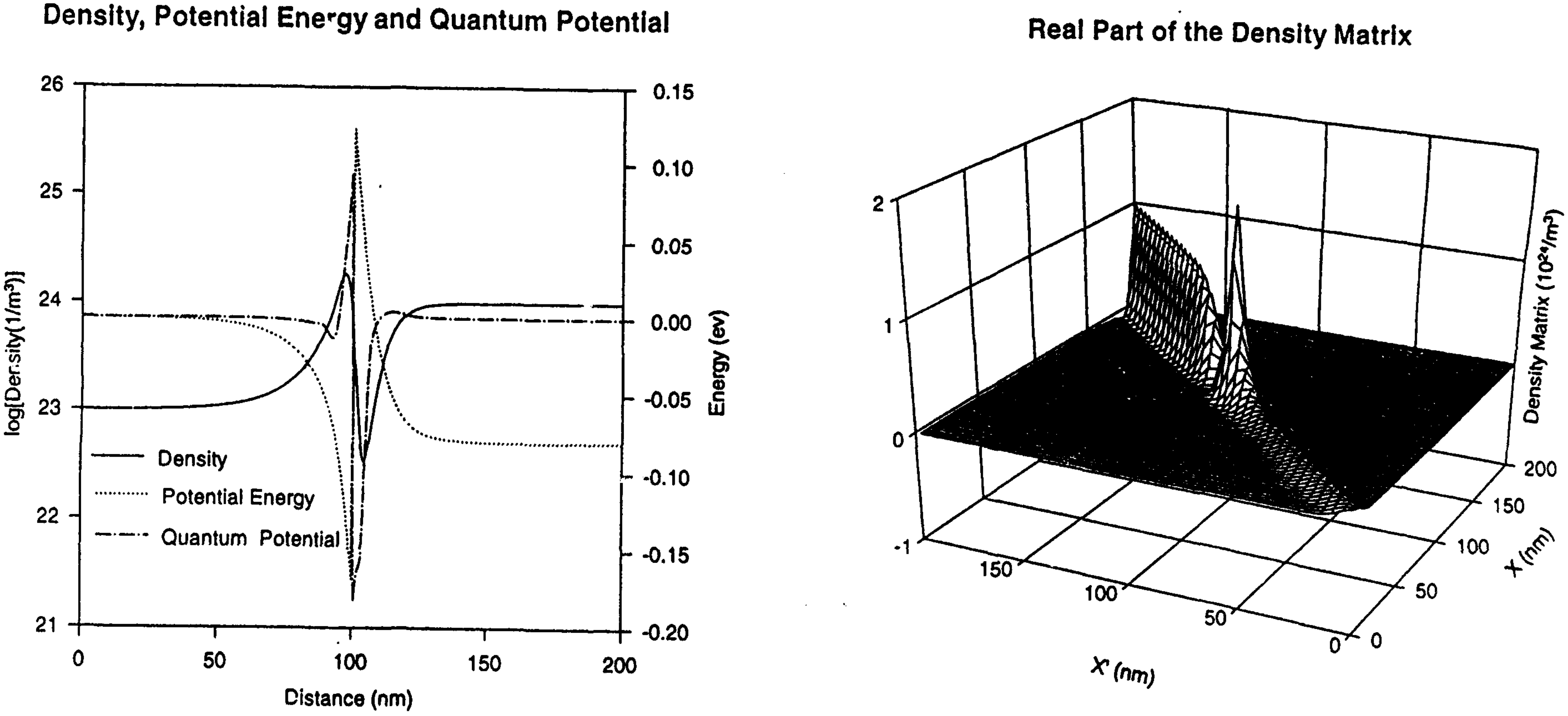
Fig. 15 Ferry & Grubin



Quantum and Potential Energy

~~[illegible]~~

Fig 11a Ferrit Grubin



Density, Potential Energy and Quantum Potential

Real Part of the Density Matrix

FIG. 17 Ferry & Grubin

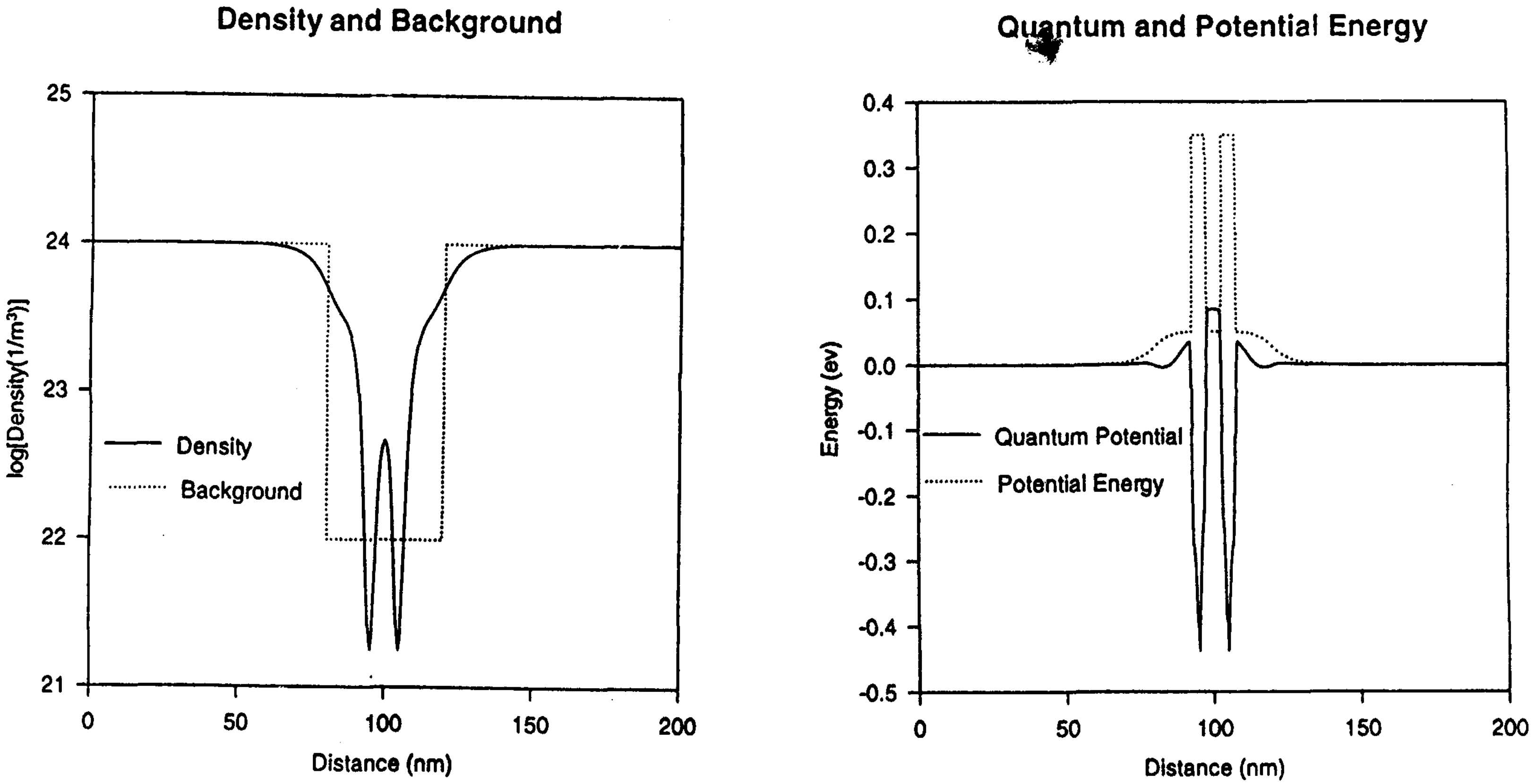
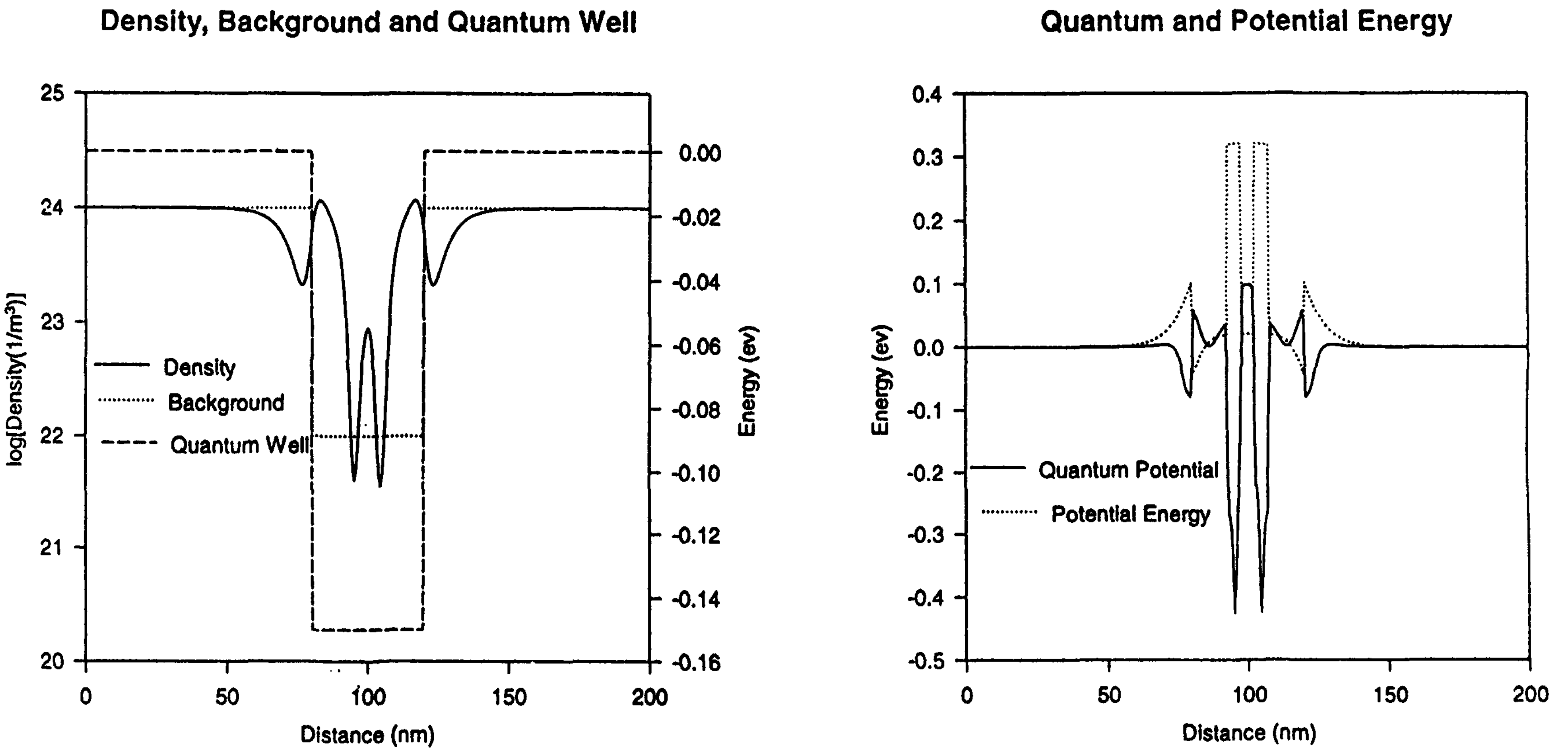


FIG. 18 Ferry & Grubin



~~Figure~~

FIG. 19 Ferry & Grubin

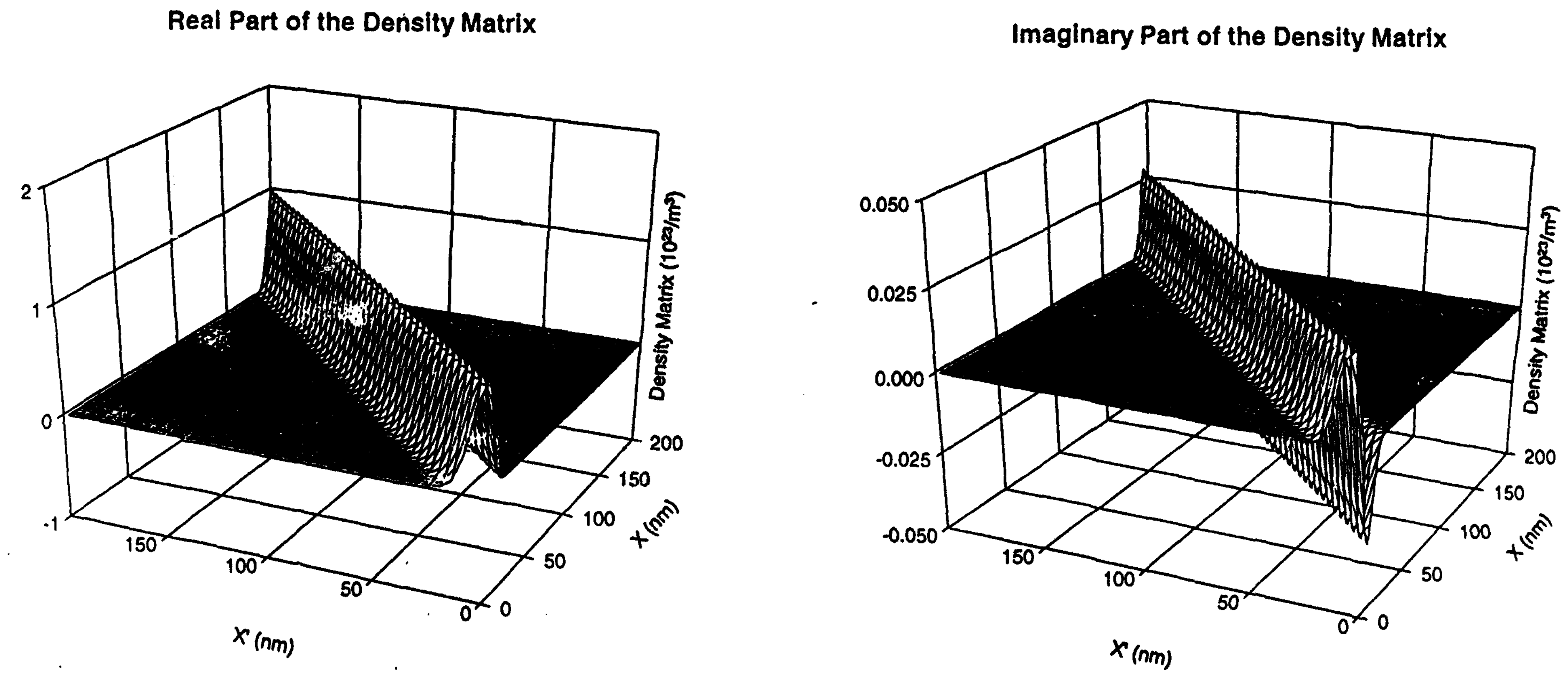
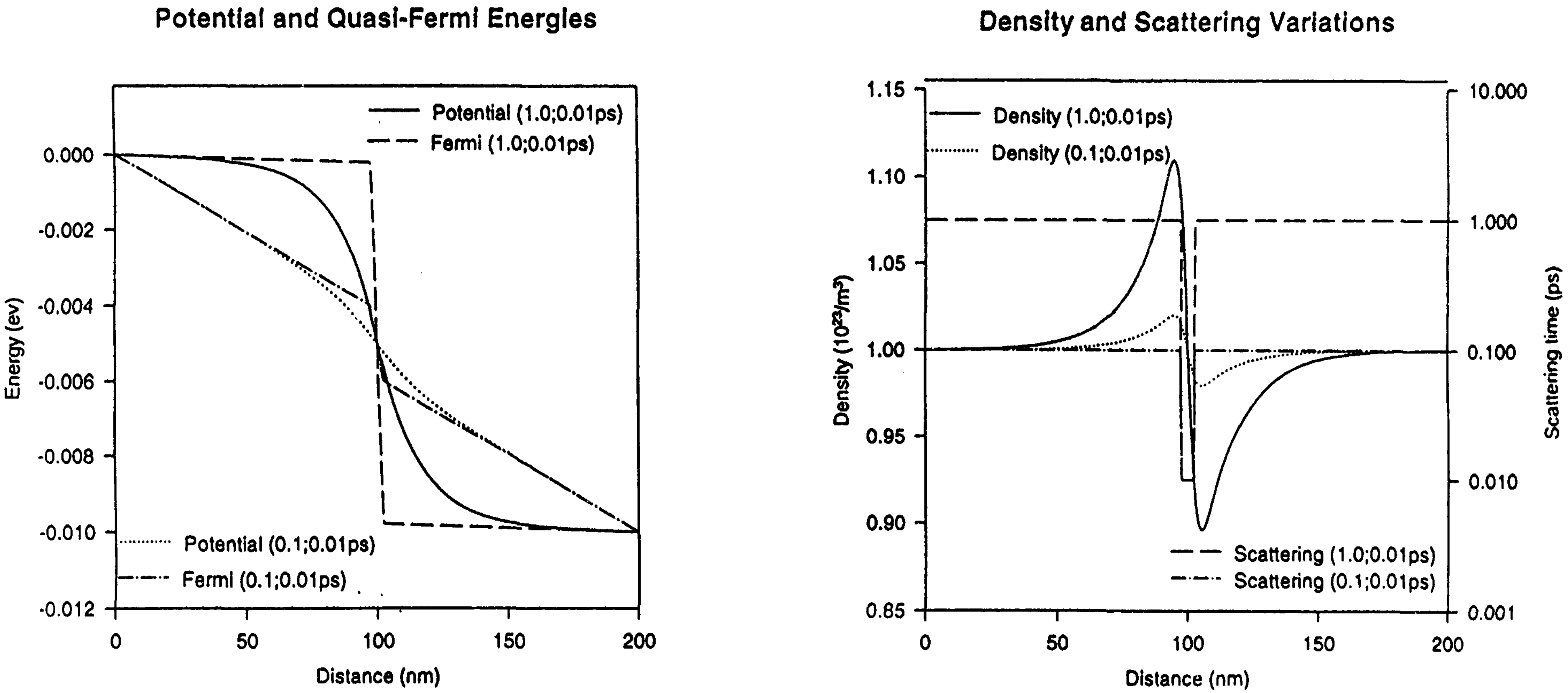
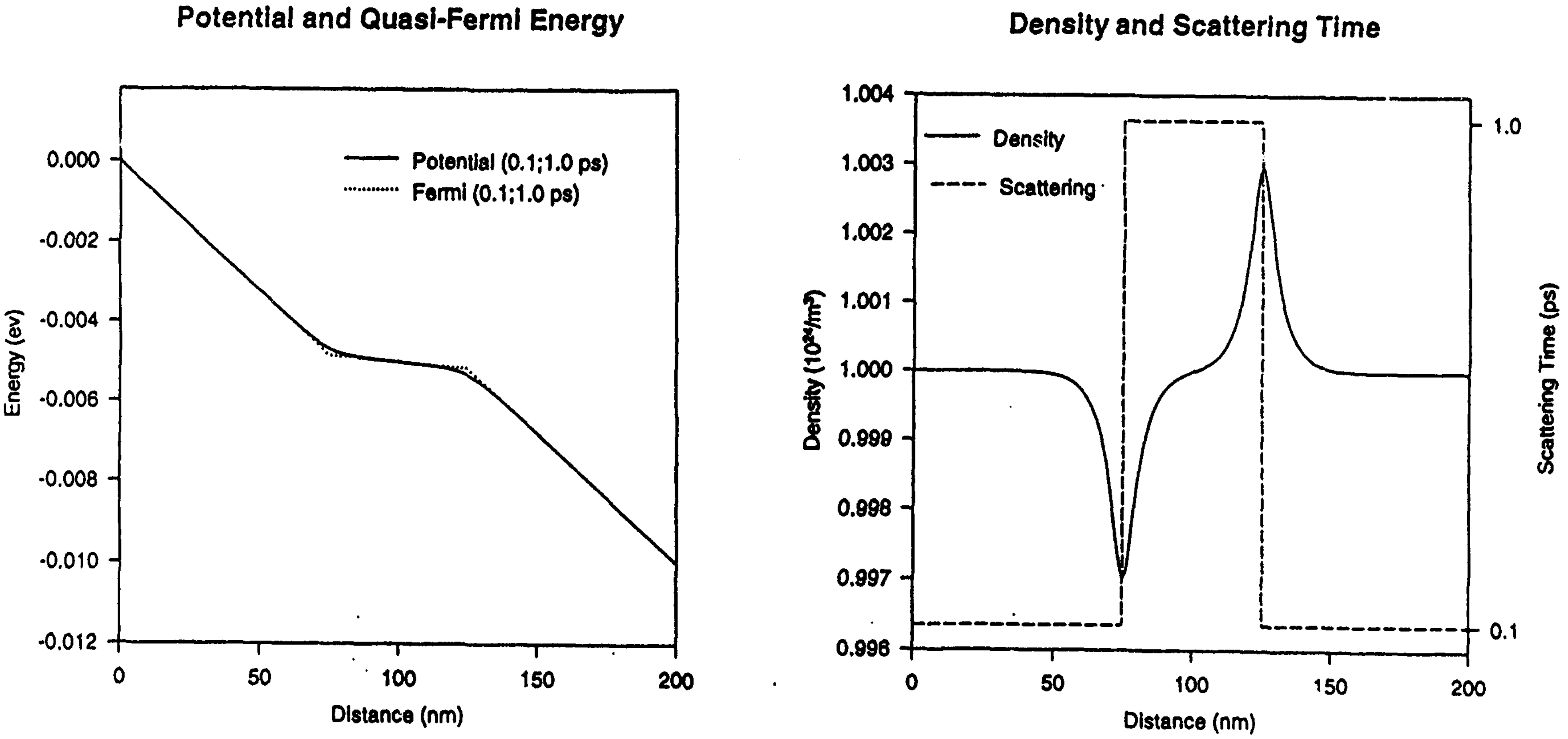


FIG. 2c Ferry & Grubin



~~Figure 12~~

FIG. 21 Ferry & Grubin

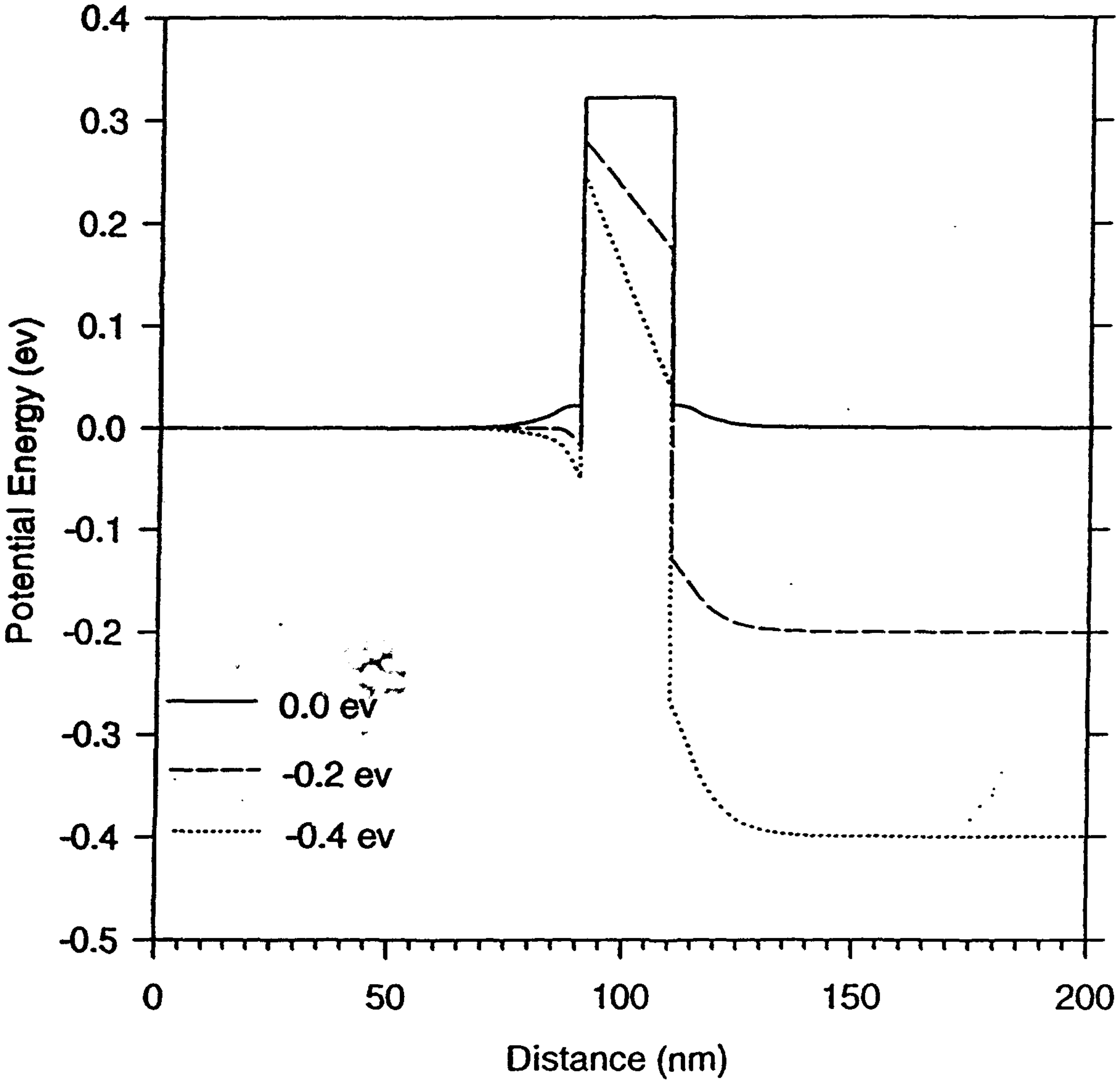

Potential and Quasi-Fermi Energy

Density and Scattering Time

~~Figure 9~~

FIG. 22 Ferry & Grubin

Potential Energy versus Distance



~~[illegible]~~

Fig. 23(a) [illegible]

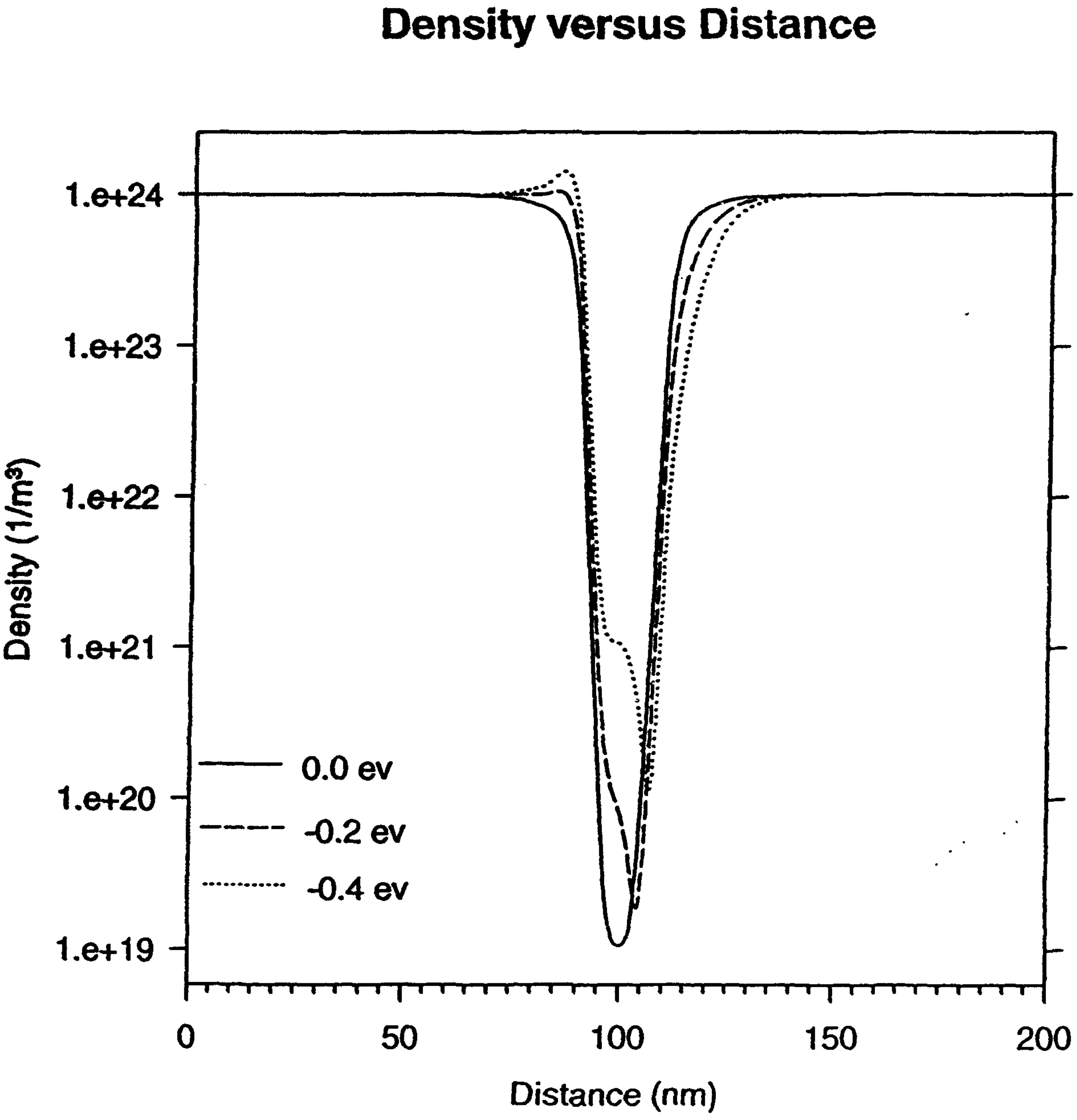
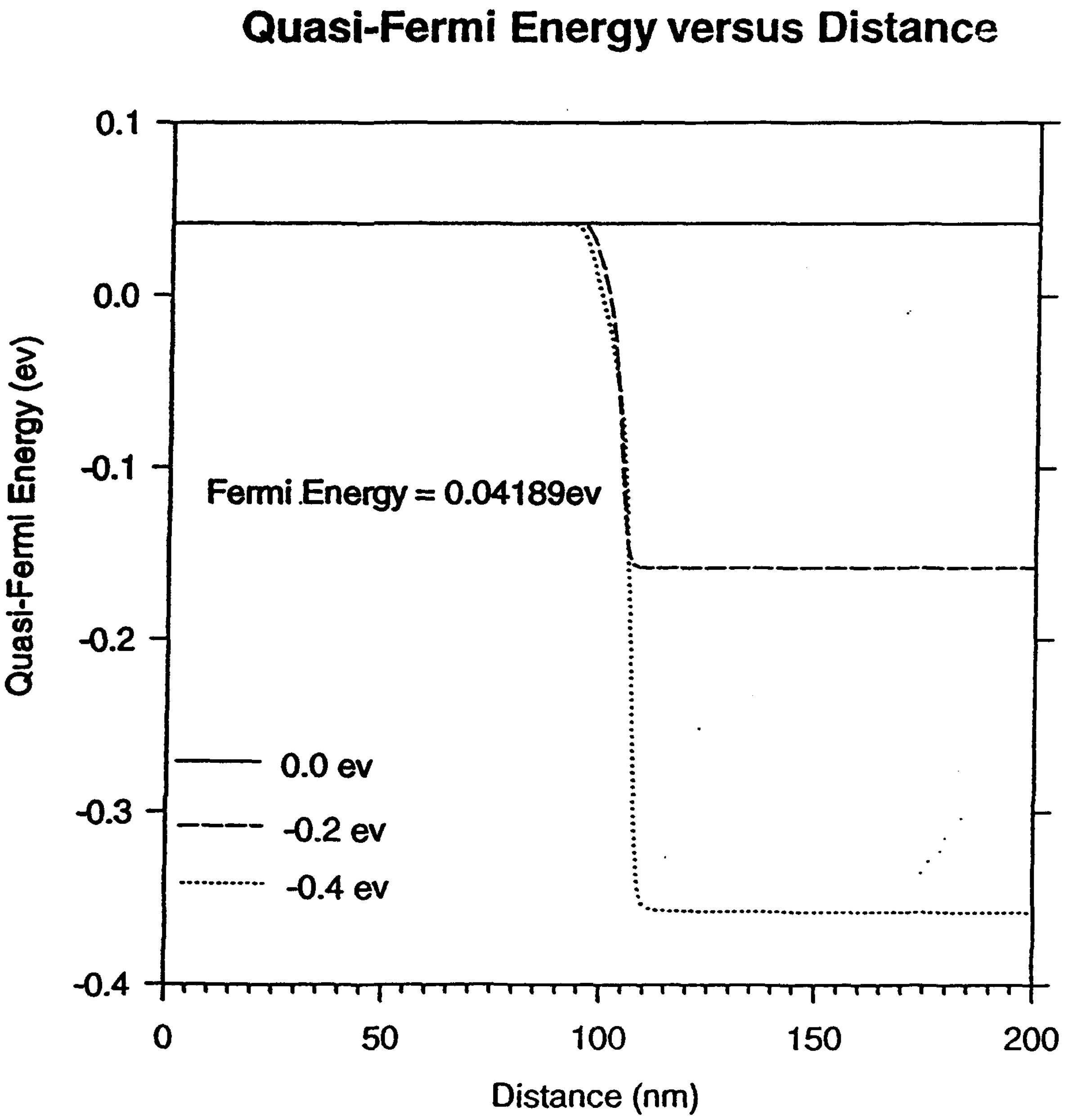


FIG. 23(b) Ferry & Grubin [illegible]



Quasi-Fermi Energy versus Distance
Quasi-Fermi Energy (ev)
0.1
0.0
-0.1
-0.2
-0.3
-0.4
Fermi Energy = 0.04189ev
0.0 ev
-0.2 ev
-0.4 ev
0
50
100
150
200
Distance (nm)

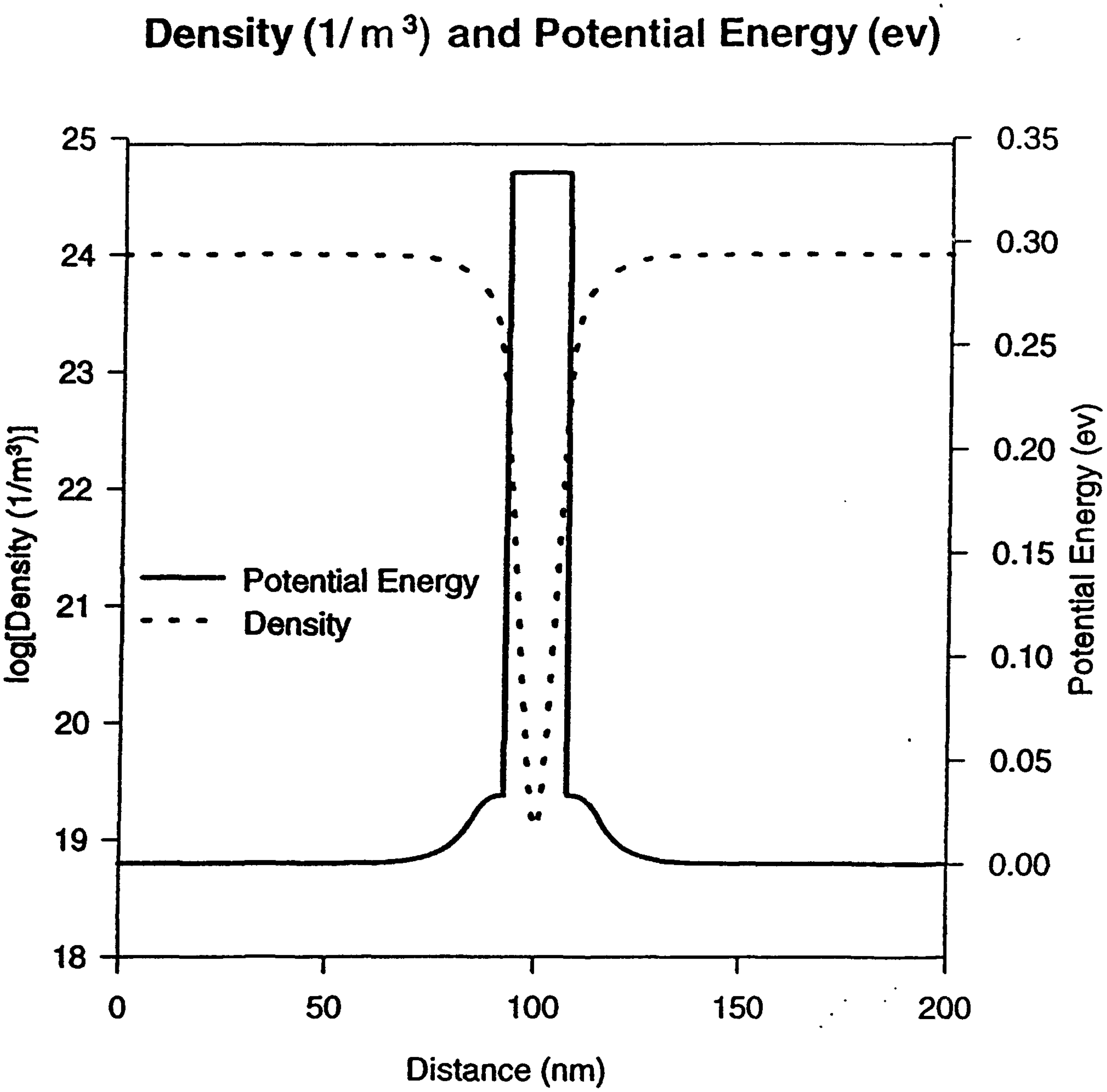
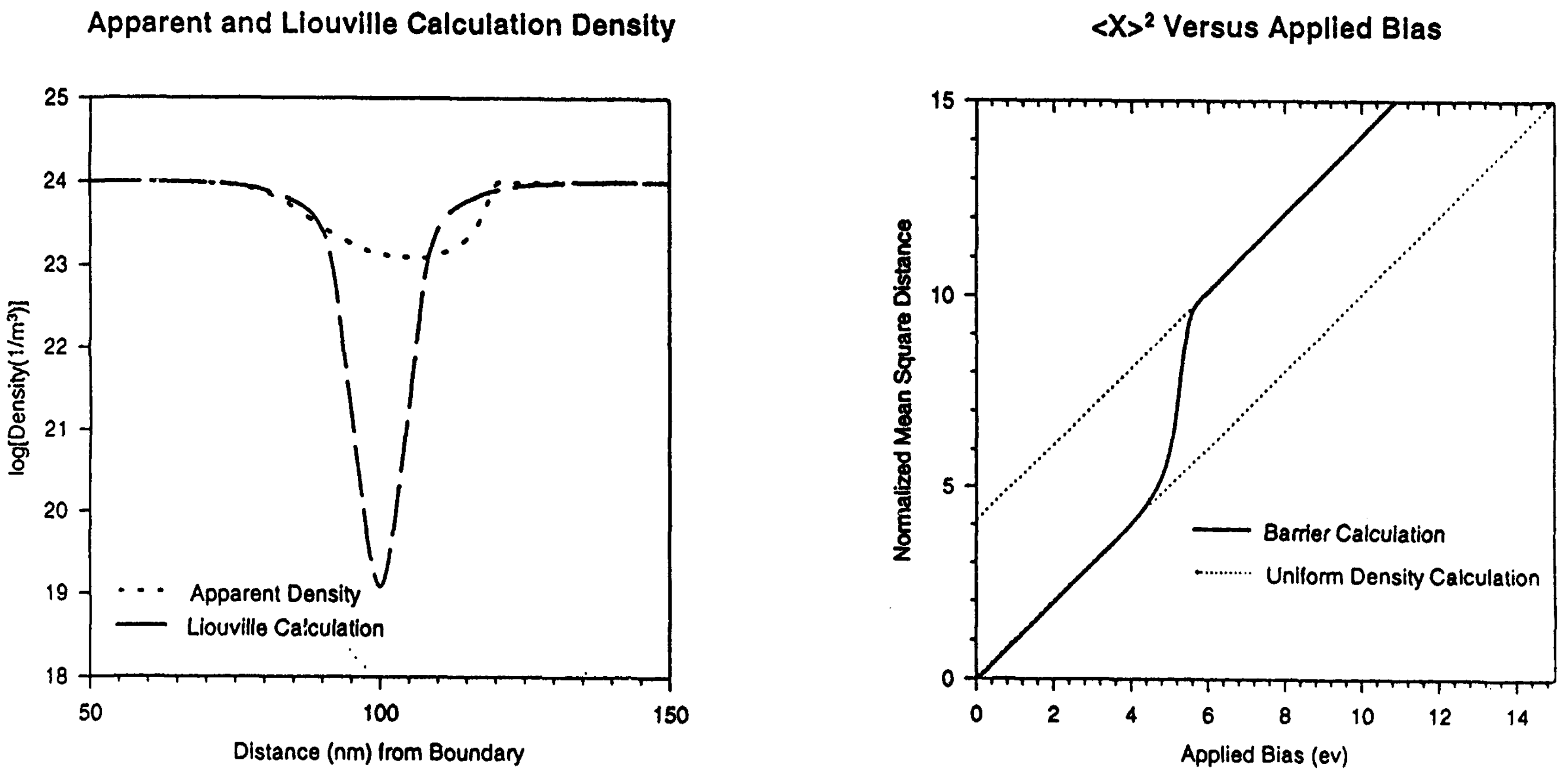


FIG. 25 Ferry & Grubin


Apparent and Liouville Calculation Density

<X>² Versus Applied Bias

~~Figure~~ X.

FIG. 26 Ferry & Grubin

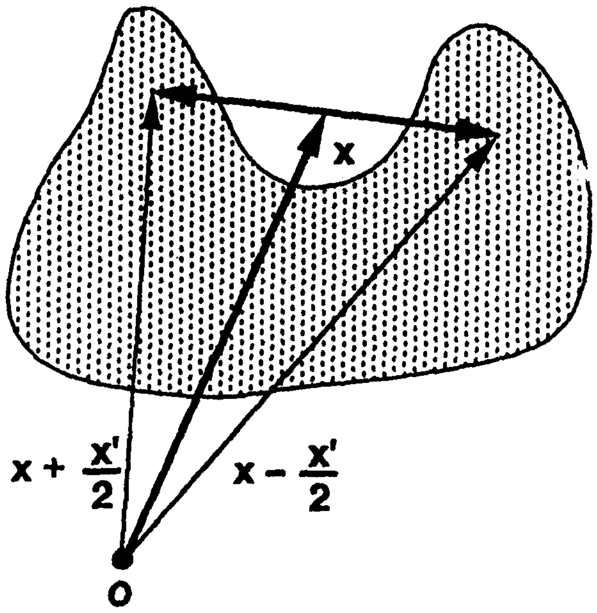

$x + \frac{x'}{2}$
$x - \frac{x'}{2}$
x
O

FIG. 27 Ferry & Grubin

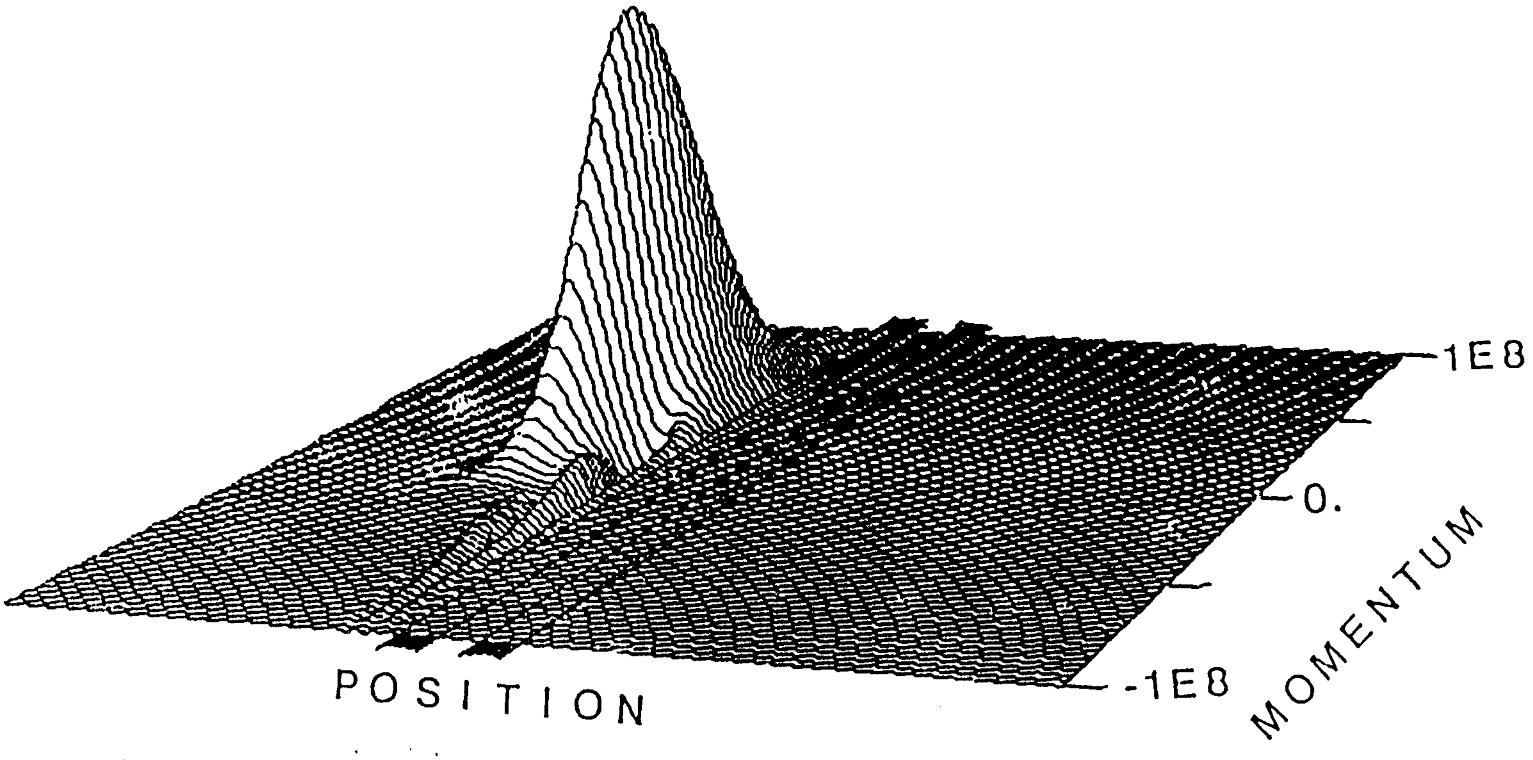
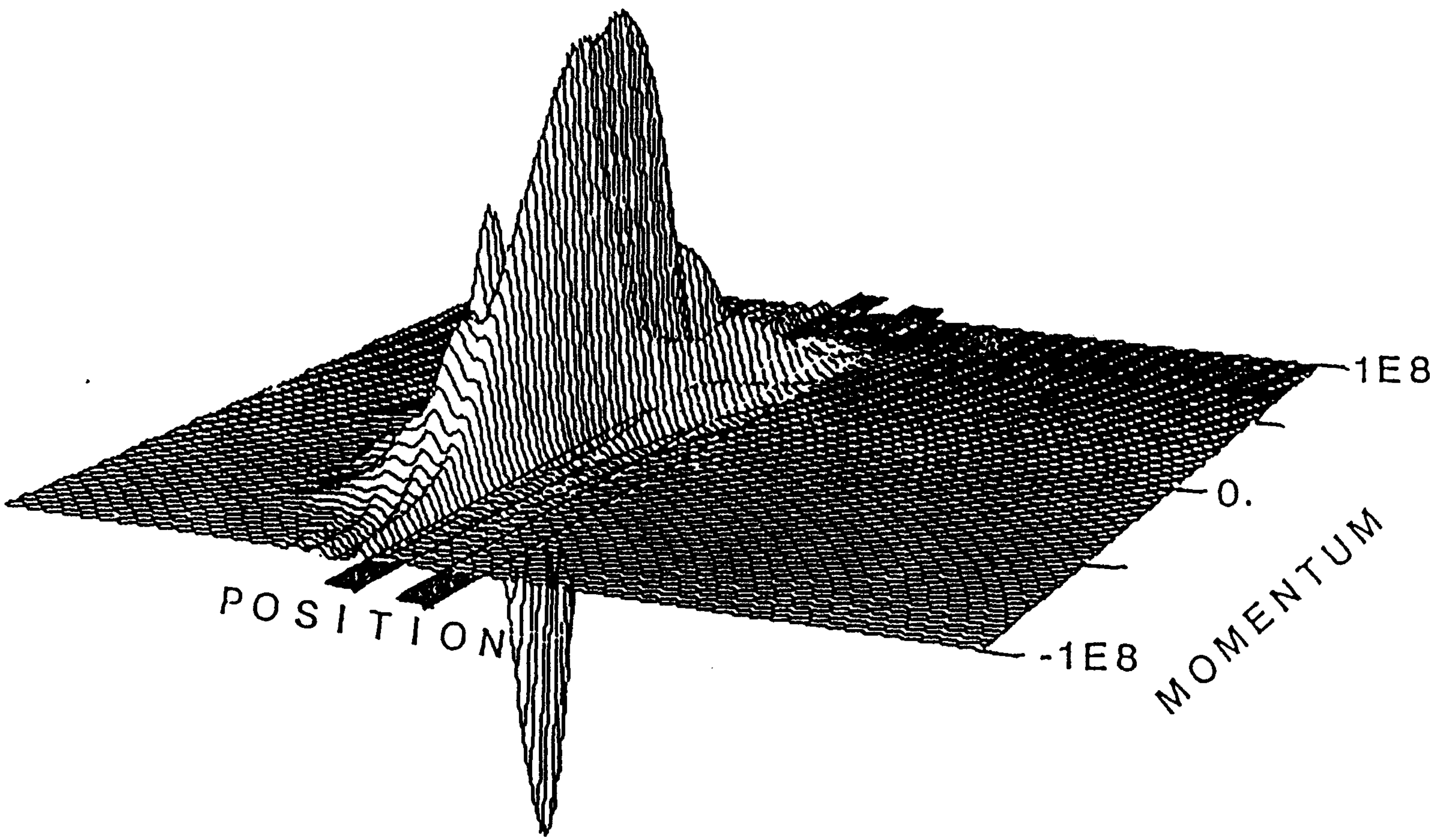


FIG. 7 3(a) Ferry & Grubin



F/6. 28(b) Ferry & Grubin

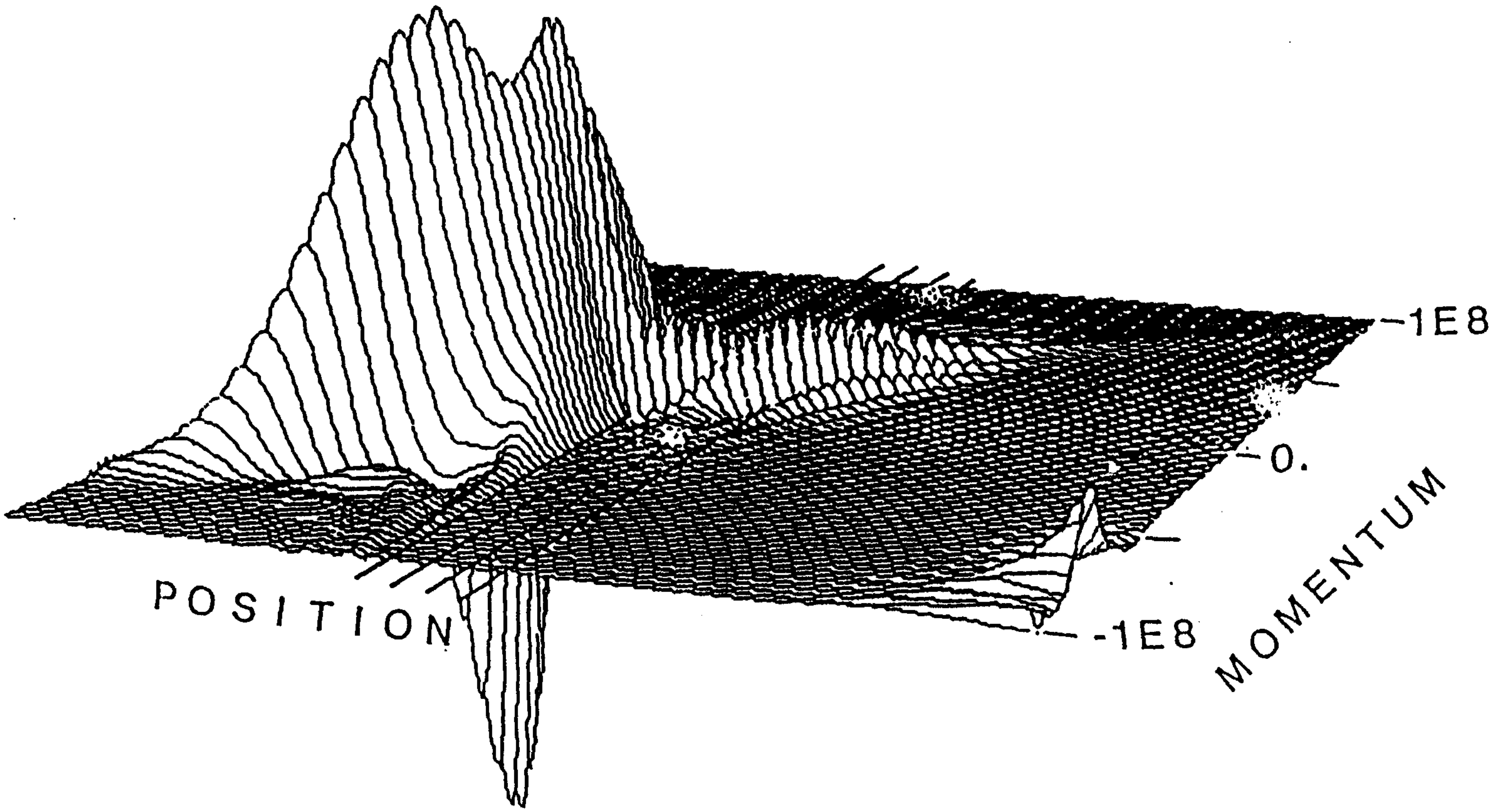


FIG. 23(c) Ferry & Grubin

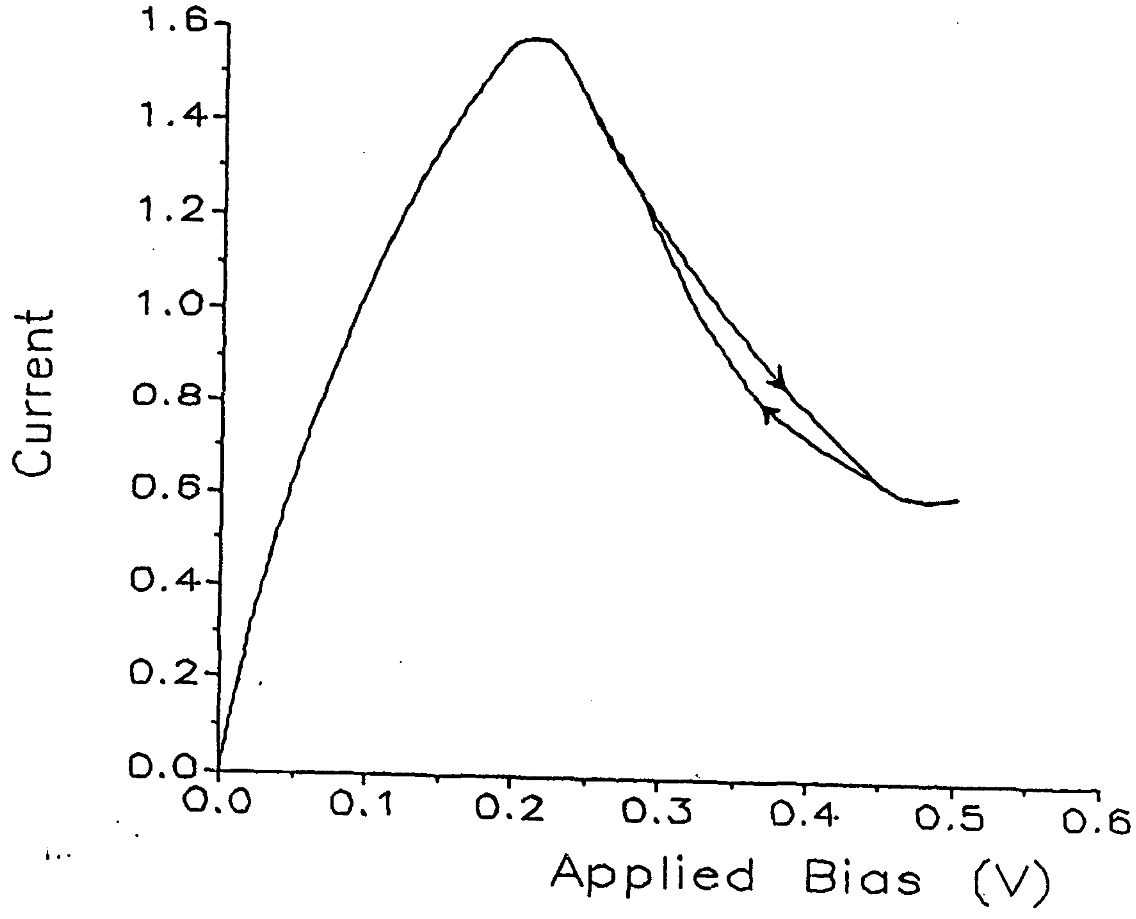


FIG. 29 Ferry & Grubin

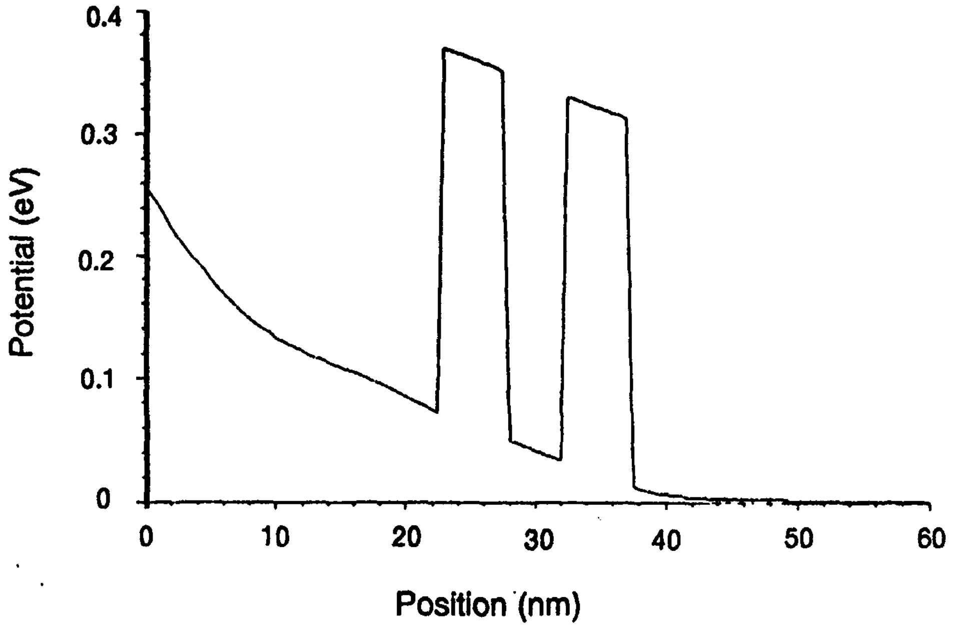


Fig. 30 Ferry & Grubin

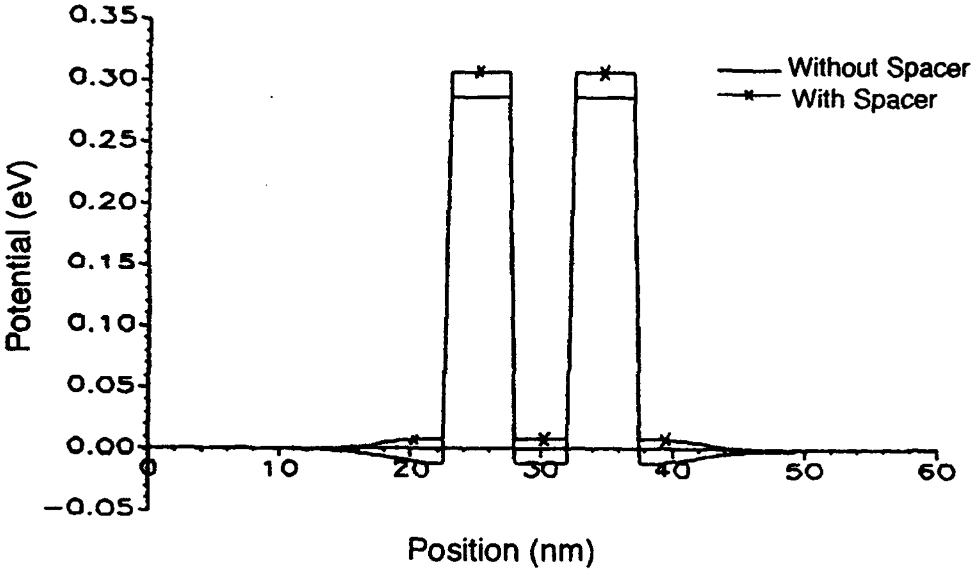


FIG. 31 Ferry & Grubin

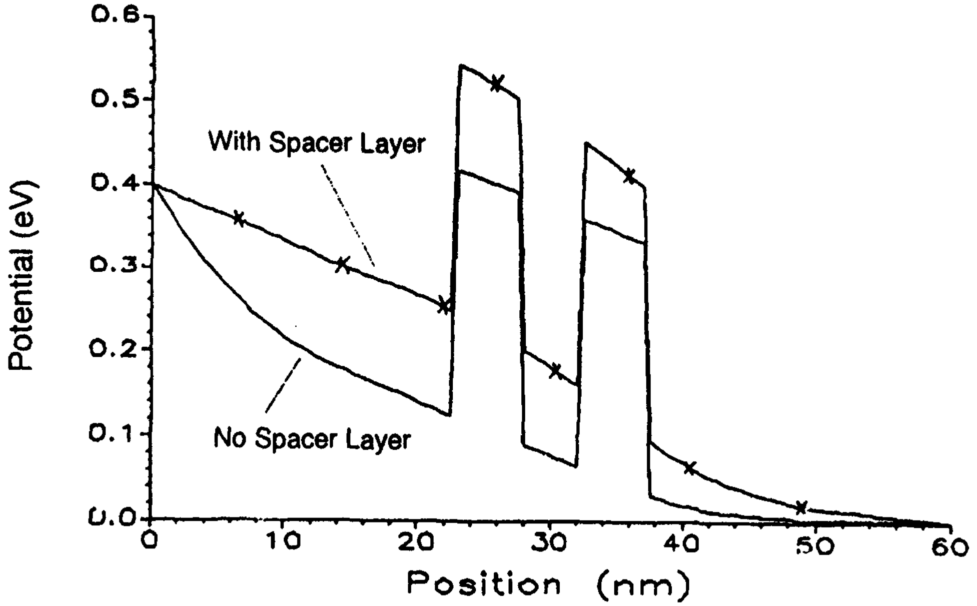


FIG. 32 Ferry & Grubin

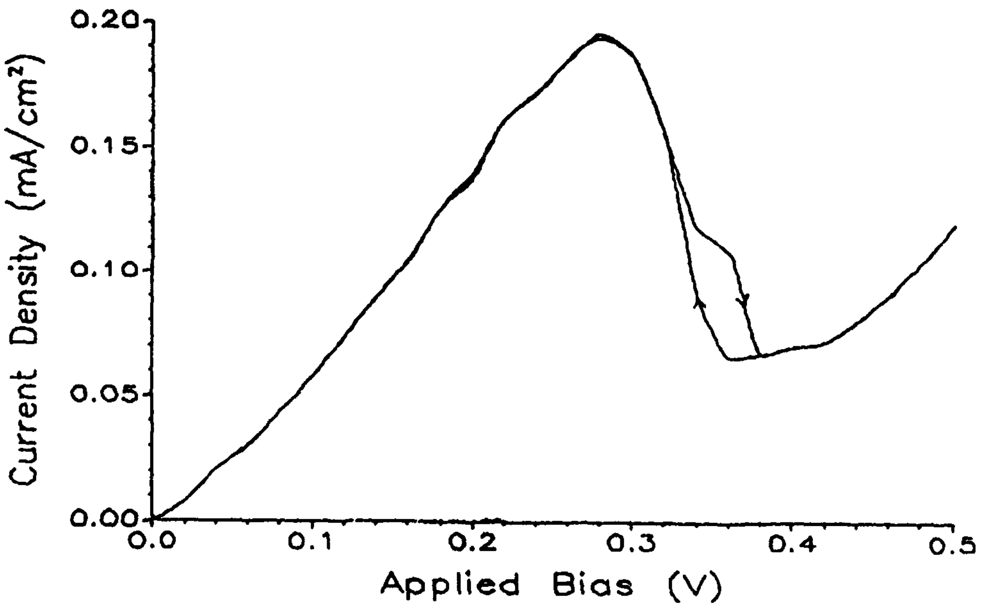


FIG. 33 Ferry & Grubin

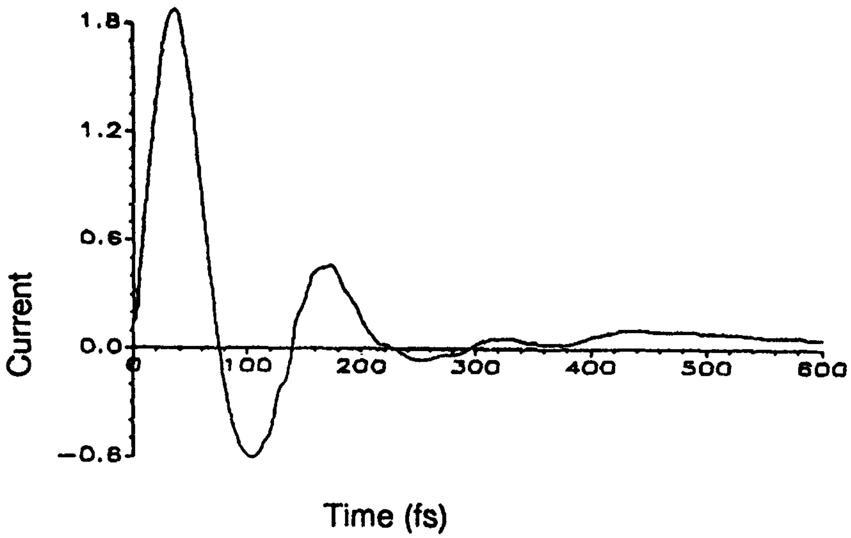


FIG-34 Ferry & Grubin

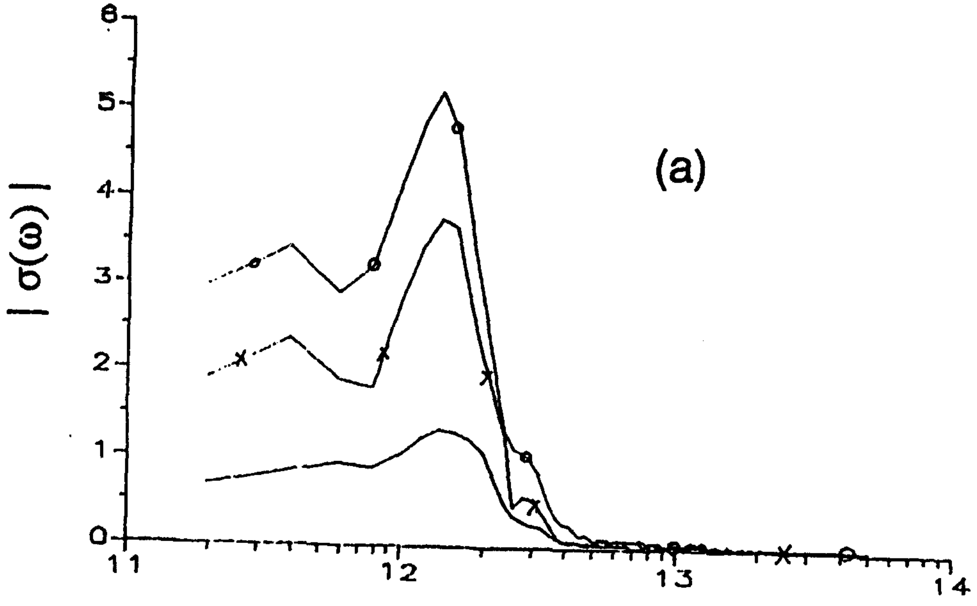


FIG. 35(a) Ferry & Grubin

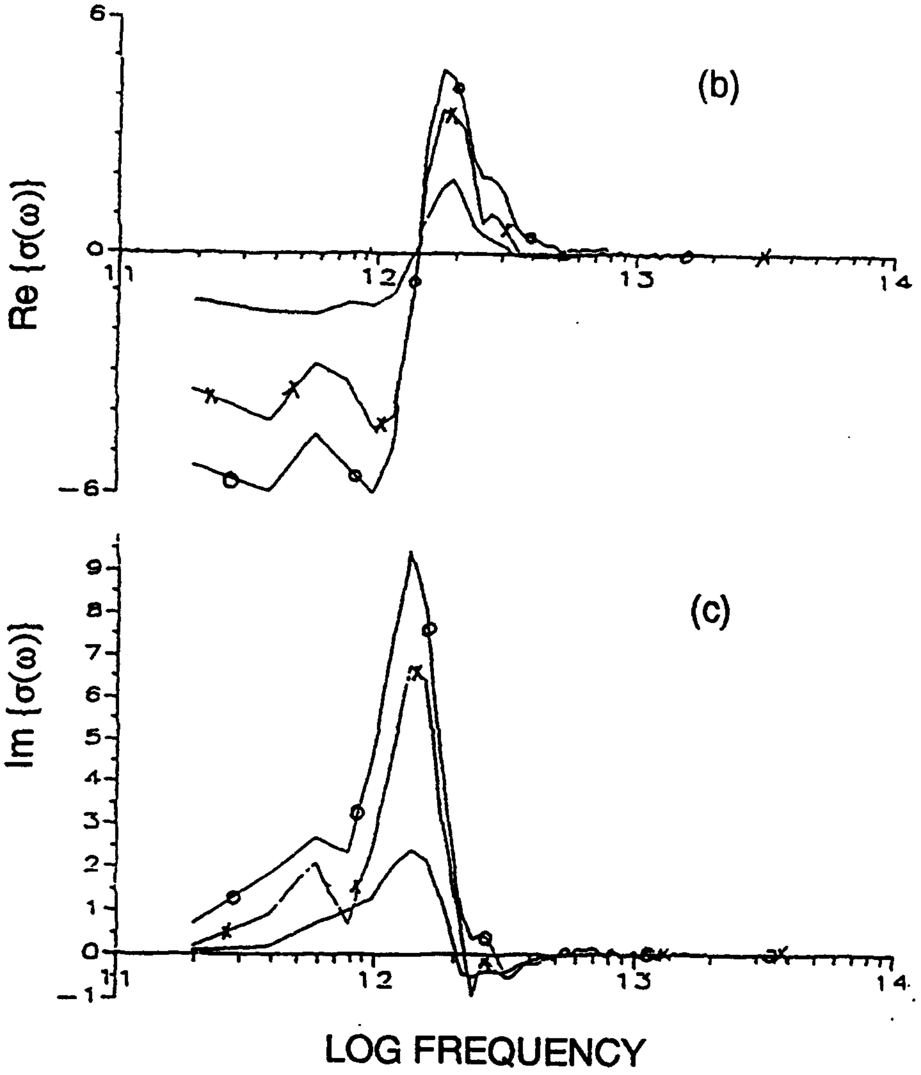

6
0
−6
(b)
Re {σ(ω)}
11
12
13
14
9
8
7
6
5
4
3
2
1
0
−1
(c)
Im {σ(ω)}
11
12
13
14
LOG FREQUENCY

Fig 5(b) & (c) [illegible]

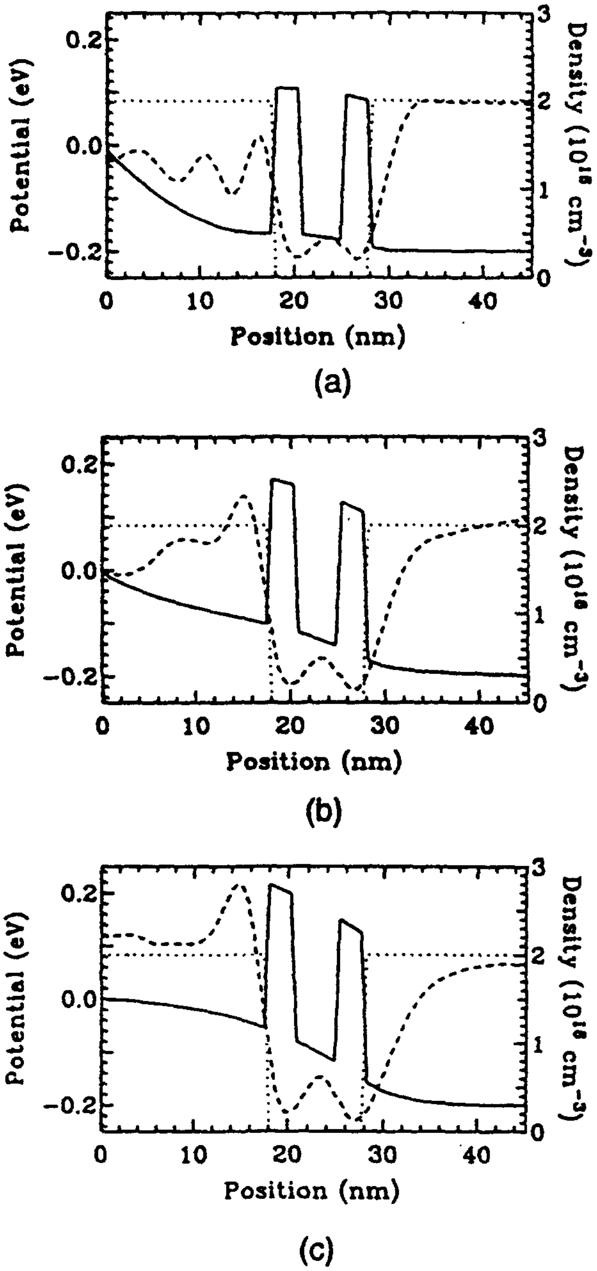
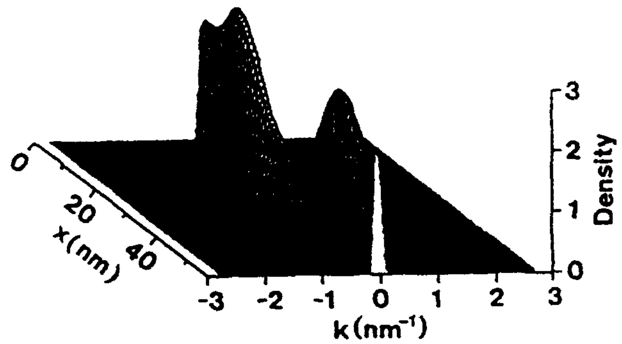
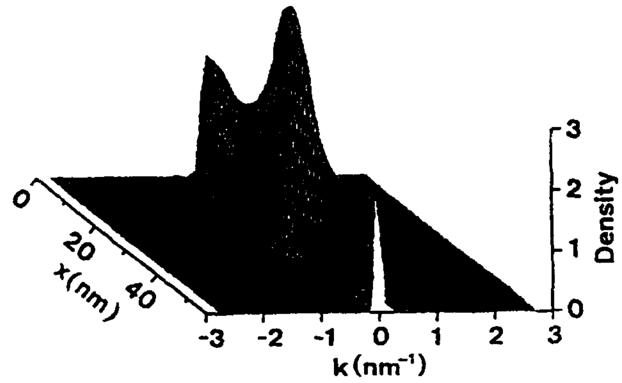


FIG 36 Ferry & Grubin

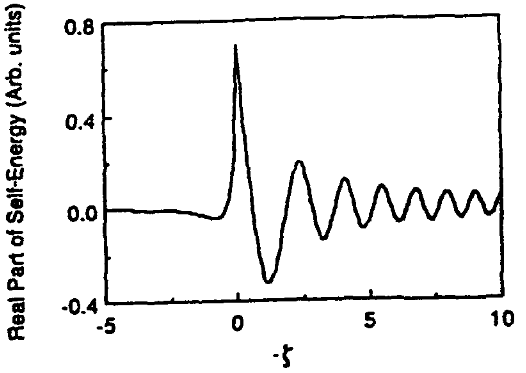


(a)

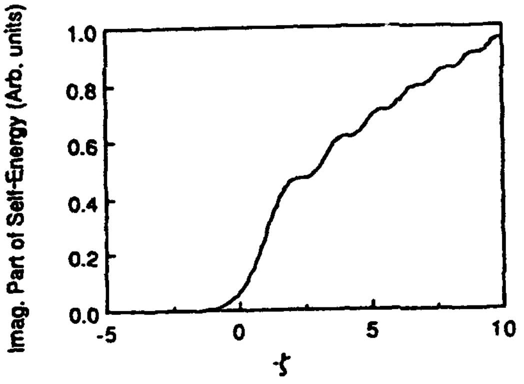


(b)

Fig 37 Ferry & Grubin


Real Part of Self-Energy (Arb. units)
0.8
0.4
0.0
-0.4
-5
0
5
10
-ζ

(a)


Imag. Part of Self-Energy (Arb. units)
1.0
0.8
0.6
0.4
0.2
0.0
-5
0
5
10
-ζ

(b)

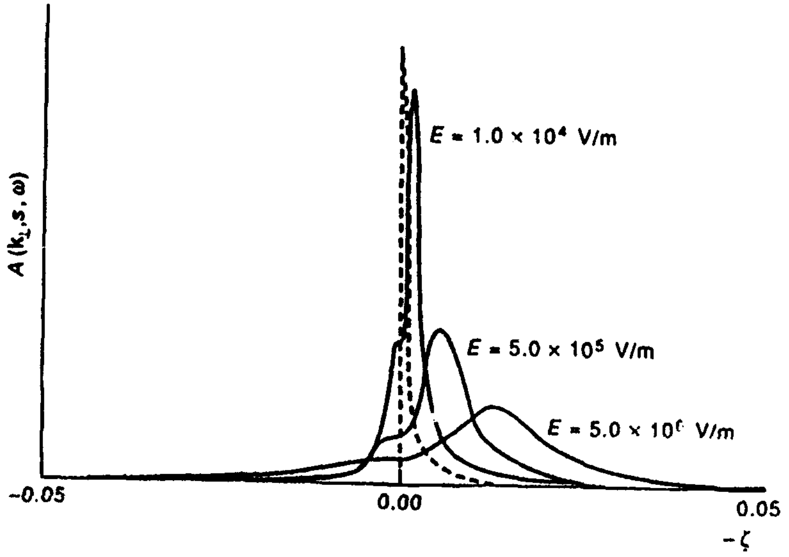


FIG. 39 Ferry & Grubin

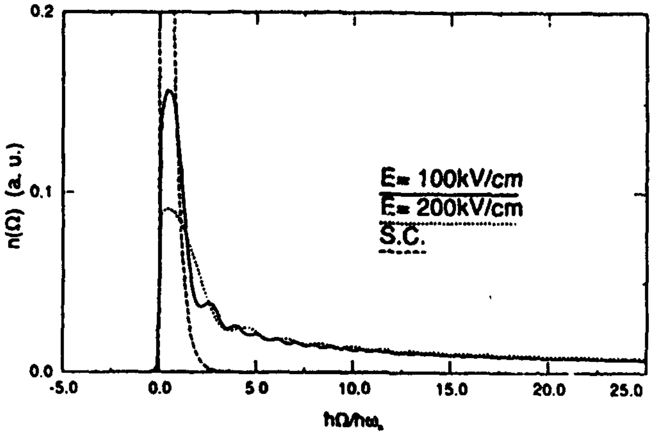


FIG. 40 Ferry & Grubin

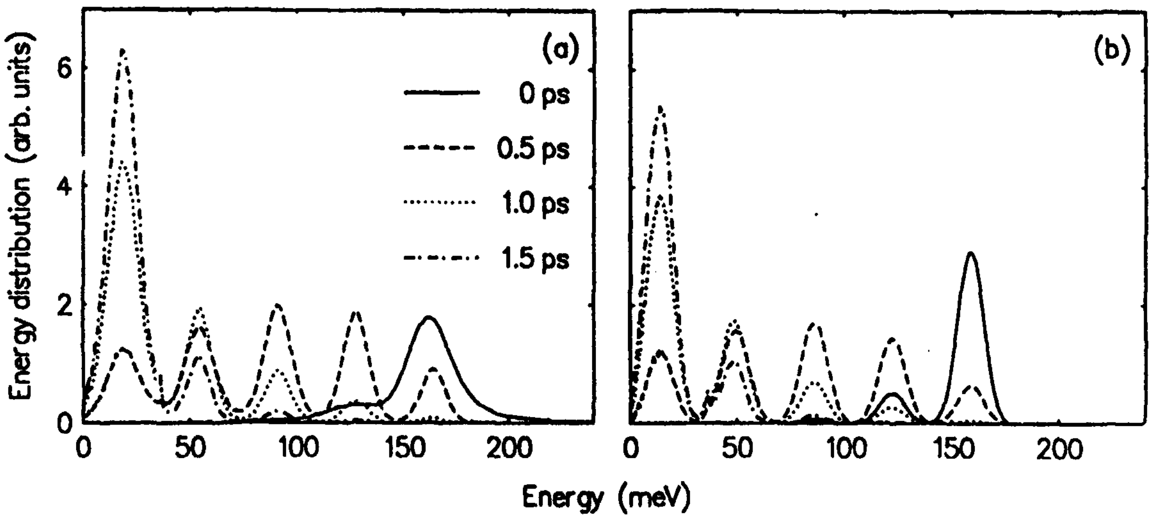


FIG. 41 Ferry & Grub.

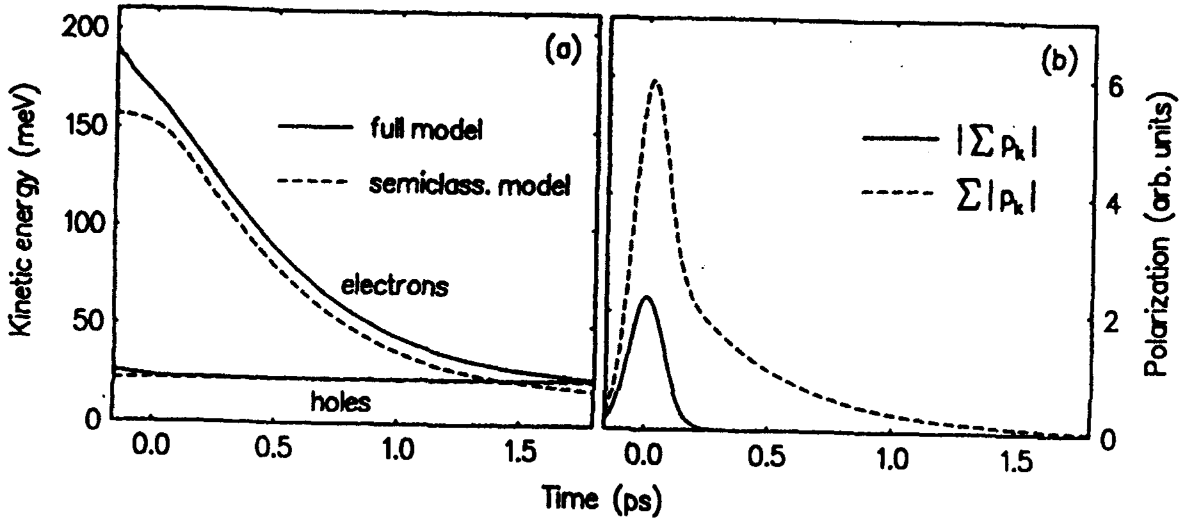


FIG. 42 Ferry & Grubin

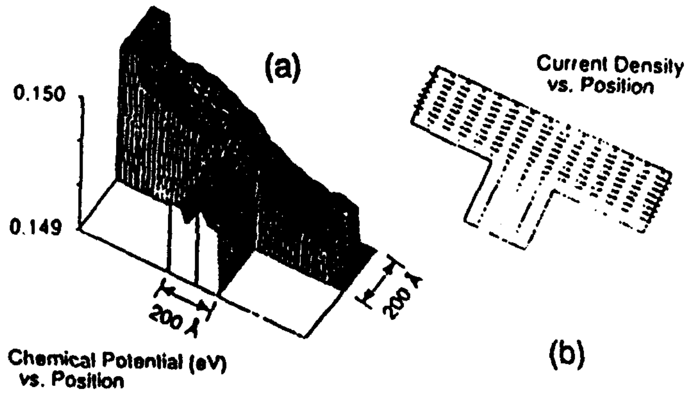
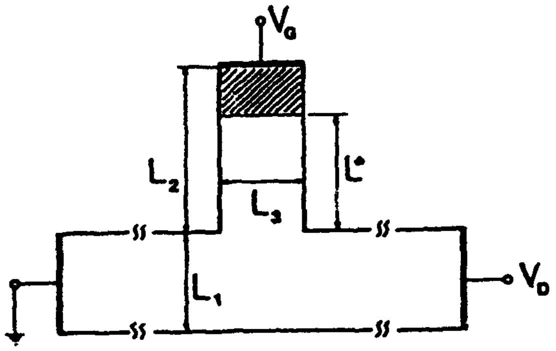
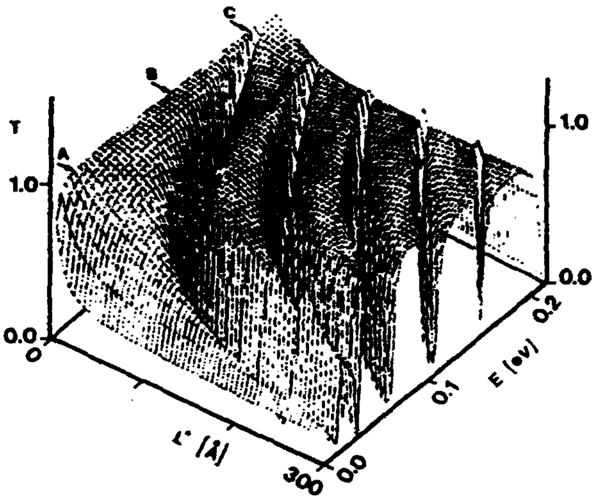


FIG. 43 Ferry & Grubin


V_G
L_2
L*
L_3
L_1
V_D

(a)



(b)

Fig. 44 Ferry & Grubin

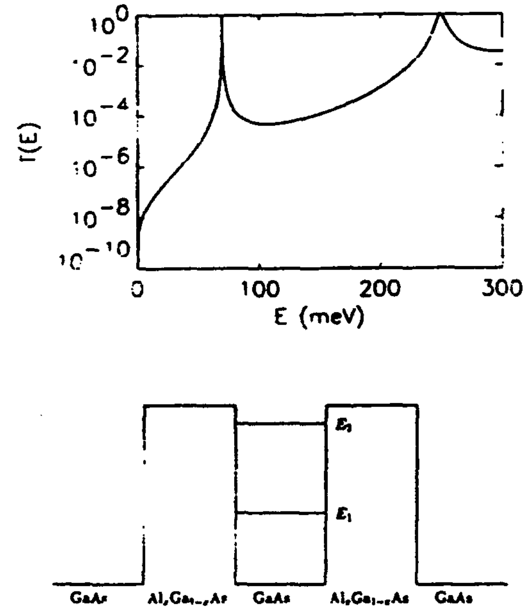


FIG. 45 Ferry & Grubin

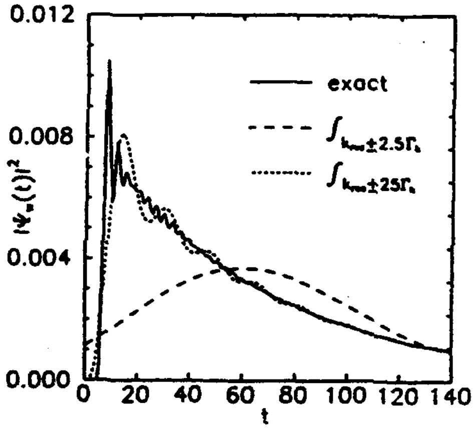


FIG. 46 Ferry & Grubin

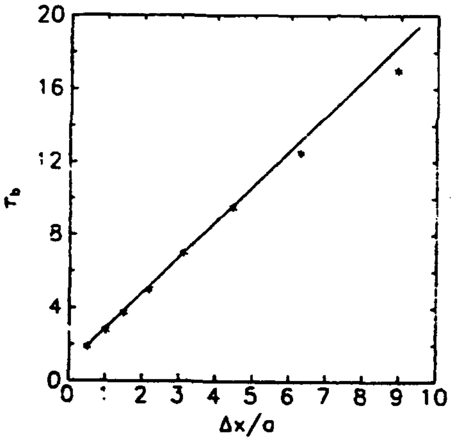


FIG-47 Ferry & Grubin

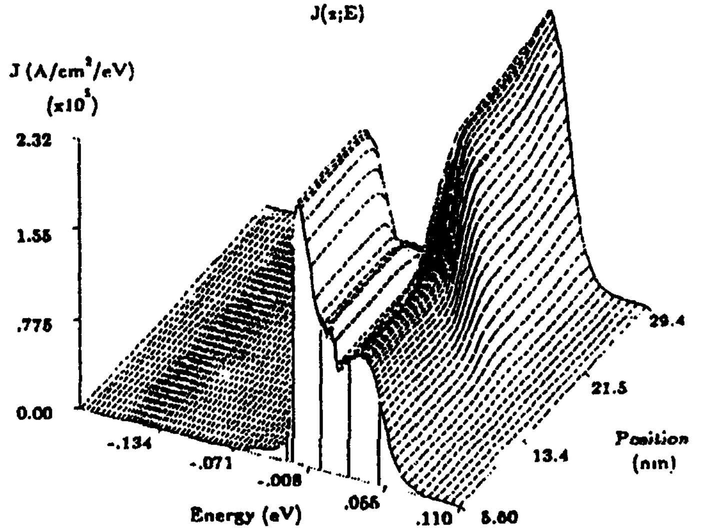


FIG. 48 Ferry & Grubin